J. R. Statist. Soc. B (1991)
53, *No.* 2, *pp.* 285–339

Procrustes Methods in the Statistical Analysis of Shape

By COLIN GOODALL†

Princeton University, USA

[*Read before* The Royal Statistical Society *at a meeting organized by the* Research Section *on Wednesday, October 17th, 1990,* Dr F. Critchley *in the Chair*]

SUMMARY

Two geometrical figures, X and Y, in $\mathbf{R}^K$, each consisting of N landmark points, have the same shape if they differ by at most a rotation, a translation and isotropic scaling. This paper presents a model-based Procrustes approach to analysing sets of shapes. With few exceptions, the metric geometry of shape spaces is quite complicated. We develop a basic understanding through the familiar QR and singular value decompositions of multivariate analysis. The strategy underlying the use of Procrustes methods is to work directly with the $N \times K$ co-ordinate matrix, while allowing for an arbitrary similarity transformation at all stages of model formulation, estimation and inference. A Gaussian model for landmark data is defined for a single population and generalized to two-sample, analysis-of-variance and regression models. Maximum likelihood estimation is by least squares superimposition of the figures; we describe generalizations of Procrustes techniques to allow non-isotropic errors at and between landmarks. Inference is based on an $N \times K$ linear *multivariate Procrustes statistic* that, in a double-rotated co-ordinate system, is a simple but singular linear transformation of the errors at landmarks. However, the *superimposition metric* used for fitting, and the *model metric*, or covariance, used for testing, may not coincide. Estimates of means are consistent for many reasonable choices of superimposition metric. The estimates are efficient (maximum likelihood estimates) when the metrics coincide. F-ratio and Hotelling's T^2-tests for shape differences in one- and two-sample data are derived from the distribution of the Procrustes statistic. The techniques are applied to the shapes associated with hydrocephaly and nutritional differences in young rats.

Keywords: GAUSSIAN MODELS; LANDMARKS; QR DECOMPOSITION; ROTATION; SHAPE SPACE; WEIGHTED PROCRUSTES ANALYSIS

1. INTRODUCTION

1.1. *Preliminaries*

The shape of a geometrical figure is commonly understood to refer to those geometrical attributes that remain unchanged when the figure is translated, rotated and scaled. A statistical approach to analysing shape is developed in this paper that applies when each geometrical figure in $\mathbf{R}^K$ consists of N labelled points and thus is represented by an $N \times K$ matrix X. Such data arise often in a biological or medical setting, when corresponding labelled points are called *landmarks* (Bookstein, 1978). Other areas in which landmark data arise include archaeology, astronomy, cartography, manufacturing, geology and geophysics. Thus in some instances landmarks may refer to the same physical markers identifiable in more than one map, satellite image,

† *Address for correspondence*: Program in Statistics and Operations Research, E-220 Engineering Quadrangle, Princeton University, Princeton, NJ 08544, USA.

 0035–9246/91/53285 $2.00

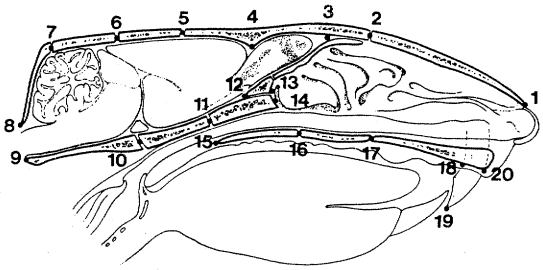

Fig. 1. Lateral section of the rat skull, showing the neural skull (left-hand half of the figure), the facial skull (right-hand half of the figure) and 20 homologous landmarks, in part at the sutures between bones

X-ray, etc. In instances of biological *homology*, landmarks are uniquely defined locations, most often in the skeleton (Fig. 1), that are identifiable across individuals and, in interspecies comparisons, linked by a presumed evolutionary pathway. At its most general, a set of landmarks is a set of labelled points, found in at least two figures, whose relative positions in the two or more figures has some scientific, educational or artistic interest to us. This abstraction allows shape theory to stand separate from issues of interpretation.

Two geometrical figures X: $N\times K$ and X': $N\times K$ are said to be congruent if they differ by a rigid body transformation (Lord and Wilson, 1984). We use *form*, or *size-and-shape*, for an equivalence class of congruent geometrical figures and loosely say that the forms X and X' are determined to within a rigid body transformation, or say that the form is the figure with location and orientation removed.

Definition. Say that two figures X: $N\times K$ and X': $N\times K$ have the same form if they are related by a rigid body transformation.

$$X' = X\Gamma + \mathbf{1}_N \boldsymbol{\gamma}^{\mathrm{T}} \tag{1.1}$$

where Γ: $K\times K$ is a rotation. $|\Gamma| = 1$, $\mathbf{1}_N$ is the N-vector of ones and $\boldsymbol{\gamma}$: $K\times 1$ is a translation. The translation–rotation pair $(\boldsymbol{\gamma}, \Gamma)$ is a transformation from the special Euclidean group acting on $\mathbf{R}^K$. We must compare the forms of X and X', instead of the figures X and X', when there is no properly defined common co-ordinate system.

Shape (Mosimann, 1970; Bookstein, 1984; Kendall, 1984, 1989) is form with size removed. A *shape* is an equivalence class of geometrical figures modulo the special similarity transformations (no reflection). Write $[X]$ for the shape of X.

Definition. Say that two figures X: $N\times K$ and X': $N\times K$ have the same shape, written $[X] = [X']$, if they are related by a special similarity transformation,

$$X' = \beta X\Gamma + \mathbf{1}_N \boldsymbol{\gamma}^{\mathrm{T}}, \tag{1.2}$$

where Γ: $K\times K$, $|\Gamma| = 1$, $\boldsymbol{\gamma}$: $K\times 1$, and $\beta > 0$ is a scalar. The triple $(\boldsymbol{\gamma}, \beta, \Gamma)$ specifies the translation, scale and rotation components of the similarity transformation from X to X'.

The focus of this paper is statistical methods for the analysis of shape. The analysis of size-and-shape is closely related. In the remainder of this section we develop both

notions, before specializing to shape (although the Procrustes methodology developed subsequently adapts very easily to size-and-shape analysis).

Kendall (1984, 1989) and Carne (1990) discuss the topological and metrical geometry of spaces for forms and shapes. The components of the similarity transformation, translation, scale and rotation, are removed in three steps. First each figure X is centred, e.g. by multiplying X on the left by an $(N-1)\times N$ Helmert matrix H. The matrix H has orthonormal rows, each orthogonal to the unit vector $\mathbf{1}_N/\sqrt{N}$. The rows of the matrix $Y=HX$ are the co-ordinates of *derived landmarks*. We say that the original figures X are in figure space, $\otimes^N\mathbf{R}^K\cong\mathbf{R}^{NK}$ and the matrices of derived landmarks are in preform space, $\otimes^{N-1}\mathbf{R}^K\cong\mathbf{R}^{(N-1)K}$. Given any probability model for X, the statistical properties of Y are easily deduced. In particular, when X is spherical Gaussian then so is Y. We need consider only scaling and rotations of Y, and can modify transformations (1.1) and (1.2) accordingly.

The next two steps are to quotient out scale and orientation, in either order (see also Goodall and Mardia (1990a, b)). As squared size statistic, r^2, we use the sum of squared deviations of landmarks about their centroid, or the sum of squares of the derived landmarks. Let $\tilde{X}=(I_N-\mathbf{1}_N\mathbf{1}_N^{\mathrm{T}}/N)X$ be the centred figure. Then

$$r=\|\tilde{X}\|=\sqrt{\mathrm{tr}(\tilde{X}^{\mathrm{T}}\tilde{X})}=\|Y\|=\sqrt{\mathrm{tr}(Y^{\mathrm{T}}Y)}. \tag{1.3}$$

To remove scale and orientation we may decompose Y uniquely as

$$\begin{aligned} Y&=rZ=rW^{\mathrm{R}}\Gamma^{\mathrm{R}}=T^{\mathrm{R}}\Gamma^{\mathrm{R}}\\ &=rZ=rW\Gamma=T\Gamma \end{aligned} \tag{1.4}$$

where $\Gamma^{\mathrm{R}}\in O(K)$ is an orthogonal matrix with $|\Gamma^{\mathrm{R}}|=\pm 1$, $\Gamma\in SO(K)$ is a special orthogonal matrix with $|\Gamma|=1$ and W^{R}, T^{R}, W and T are lower triangular matrices with non-negative diagonal elements, *apart from* W_{KK} and T_{KK} which may be negative, positive or zero. $\|T^{\mathrm{R}}\|=\|T\|=r$ and $\|Z\|=\|W^{\mathrm{R}}\|=\|W\|=1$, and $[X]=[Y]=[T]=[W]$.

The decomposition $Y=T^{\mathrm{R}}\Gamma^{\mathrm{R}}=T\Gamma$ is the *polar*, or *QR*, decomposition of Y, and T^{R} is said to contain the *rectangular co-ordinates* of Y. The two alternatives occur in computation according to whether Householder reflections or only Givens rotations are used in the *QR* algorithm (Golub and Van Loan, 1983). In multivariate analysis the orthogonal group is used. In shape theory the distinction is important: either reflection (denoted by the superscript R) is allowed, or it is not. We are led to two additional definitions, parallel to definitions 1 and 2, for what may be called *reflection form* (*reflection size-and-shape*) and *reflection shape*.

Expressions (1.4) provide co-ordinates for shape and form, and reflection shape and reflection form. The structural non-zeros of T^{R} are reflection size-and-shape co-ordinates. The $NK-K(K+1)/2$ structural non-zeros of the *form matrix* T are size-and-shape co-ordinates, in the space of forms $S\Sigma_K^N$ (Kendall's notation). Furthermore Z is a point on the sphere of preshapes $P\Sigma_K^N\equiv S_K^N\cong S^{(N-1)K-1}$. The structural non-zeros of the *shape matrix* W are *direction co-ordinates* for shape, topologically in S^m, where $m=NK-K(K+1)/2-1$. We may then introduce m generalized polar shape co-ordinates. For the shapes of planar triangles $\Sigma_2^3\cong S^2(\frac{1}{2})$,

$$W=\begin{pmatrix} \cos\theta_1 & 0\\ \sin\theta_1\cos\theta_2 & \sin\theta_1\sin\theta_2 \end{pmatrix}, \tag{1.5}$$

and $(2\theta_1, \theta_2)$ are the usual spherical polar co-ordinates, colatitude and longitude.

The space of reflection shapes (reflection forms) involves a 2:1 identification of shape (form) space. We introduce the notation $R\Sigma_K^N$ and $SR\Sigma_K^N$ for reflection shape space and reflection form space respectively. Since a rotation followed by reflection in $\mathbf{R}^K$ is a rotation when $\mathbf{R}^K$ is embedded in $\mathbf{R}^{K+1}$, each reflection space is the subspace of Σ_{K+1}^N or $S\Sigma_{K+1}^N$ containing rank deficient shapes.

Kendall (1984, 1989), Carne (1990) and Le (1990a) characterize the geometry of shape space both topologically and metrically. Generalizing the result that shape space for planar triangles is the sphere, $\Sigma_2^3 = S^2(\frac{1}{2})$, with radius $\frac{1}{2}$, Kendall (1984) shows that shape space for more complicated planar configurations is complex projective space $\Sigma_2^N \cong \mathbf{C}P^{N-2}(4)$, with so-called sectional curvature 4. Kendall and Le (1990) characterize the geometry of Σ_K^N with N and K arbitrary.

The Euclidean geometry of figure space induces Euclidean geometry in preform space (H is linear), and spherical geometry in preshape space $P\Sigma_K^N = S_K^N \cong S^{(N-1)K}$. The preshape sphere is partitioned into *fibres* by the rotation group $SO(K)$. Two preshapes on a given fibre differ by a rotation. The fibre through Z is $\{Z\Gamma\colon \Gamma \in SO(K)\}$, and the fibre is the *orbit* of Z under the action of $SO(K)$. Fibres in S_K^N correspond one to one with points in shape space Σ_K^N. The squared Procrustes distance between $[Z_1]$ and $[Z_2]$, $d^2([Z_1], [Z_2])$, is the smallest Euclidean distance between any pair of preshapes on the corresponding fibres,

$$d^2 = \min_{\Gamma \in SO(K)} \| Z_1 - Z_2\Gamma \|^2 = 2\{1 - \max_{\Gamma \in SO(K)} (Z_1^{\mathrm{T}} Z_2 \Gamma)\}. \tag{1.6}$$

Some researchers (Kendall, Carne) prefer to work with Riemannian or geodesic distance α, which on the sphere is just the great circle distance, and is related to Euclidean distance by $d = 2\sin(\alpha/2)$ (in both preshape and shape space). The quotient mapping from S_K^N to Σ_K^N is called a Riemannian submersion, because it preserves Riemannian distances.

These formulae and discussion reveal the parallels between spherical statistics and shape theory. Shape theory involves an additional layer of complication, through the rotations, and preshape space S_K^N has more structure (e.g. the fibres) than the sphere $S^{(N-1)K-1}$. Spherical statistics and Procrustes methods in shape theory (and preshape theory) both use the chord distance d as a convenient and pervasive alternative to Riemannian distance. For planar triangles equation (1.6) can be avoided, since $[Z]$ is a point on $S^2(\frac{1}{2})$ and d is the chord distance. This approach has been developed by several researchers, including Kendall (1985, 1989), Mardia (1989a) and Goodall *et al.* (1990).

Recalling the QR decomposition (1.4), the distance between any two shape matrices W_1 and W_2, or form matrices T_1 and T_2, is not a simple function of (W_1, W_2) or (T_1, T_2) but is instead obtained by replacing the Zs by Ws or Ts in equation (1.6) and finding the minimizing Γ. Thus equation (1.6) provides a natural metric for lower triangular matrices. The theory of size-and-shape (almost) addresses the geometrical and statistical properties of the lower triangular matrix of the QR decomposition. The theory of shape (almost) addresses the properties of the lower triangular matrix scaled to unit size. The niggledy 'almost' is a consequence of excluding reflections and is (hopefully) usually unimportant in practice. Shape theory appears to be the first systematic investigation of matrices modulo rotations, instead of modulo orthogonal transformations.

1.2. *Euclidean Co-ordinates for Shapes*

Let $x_1, x_2, \ldots, x_N \in \mathbf{C}$ be the complex co-ordinates of the landmarks of a planar figure, $K=2$. Let $\mathbf{u}=(u_i)$, where

$$u_i = \frac{x_{i+2}-x_1}{x_2-x_1}, \qquad i=1, \ldots, N-2. \tag{1.7}$$

The u_i comprise $2(N-2)$ shape co-ordinates, called *Bookstein co-ordinates* on account of Bookstein (1984, 1986, 1991). The Euclidean geometry of $\mathbf{u} \in \mathbf{R}^{2(N-2)}$ is a useful approximation to the metric geometry of $\Sigma_2^N \cong \mathbf{C}P^{N-2}(4)$, allowing concentrated shape data to be analysed as an approximately multivariate Gaussian sample. For triangles, $N=3$, Bookstein co-ordinates are a stereographic projection of the spherical co-ordinates. Bookstein also develops a distinct global hyperbolic geometry for $\mathbf{u}$ (cf. Small (1988)). Numerous researchers use the u_i (Bookstein and Reyment, 1989; Kendall, 1984, 1985; Goodall, 1990a; Mardia and Dryden, 1989a, b; Mardia, 1989a, b; Goodall *et al.*, 1990; Small, 1984; Watson, 1986).

Kendall's (1984) *co-ordinates* are the ratios $u_i^* = y_{i+1}/y_1$ of the complex co-ordinates y_i of the derived landmarks $Y=HX$. The u_i^* comprise the rectangular part of the QR decomposition of Y when scale is $\|y_1\|$, so the first row of the shape matrix W can be ignored. Analogously, the u_i comprise the rectangular part of the QR decomposition when the Helmert matrix H is replaced by a 'difference' matrix, $(-\mathbf{1}_{N-1} | I_{N-1})$, and scale is $\|x_2-x_1\|$.

Both Bookstein and Kendall co-ordinates generalize to arbitrary dimension by this approach (Goodall and Mardia, 1990a). The use of complex co-ordinates for landmarks of planar figures is an elegant alternative (Kendall, 1984; Kent, 1990; Bookstein, 1991) to the more general matrix methods used here.

1.3. *Overview*

The organization of the paper is as follows. A summary of the notation used in subsequent sections is given in Table 1. Procrustes estimation of the mean shape, and shape differences, is discussed in Sections 2 and 3 from a geometrical standpoint. Section 4 introduces the singular value decomposition (SVD) and isometries of shape space. Gaussian models for landmark data are defined in Section 5 for general error covariances, and maximum likelihood estimation is discussed. The one- and two-sample models are given as examples in Section 6. Conventional unweighted Procrustes algorithms are extended in Section 7 (ordinary) and Section 8 (generalized) to the weighted case. However, the metric used for estimation need not coincide with the model covariance (Section 9). Iterative algorithms leading to maximum likelihood estimates (MLEs) for the one- and two-sample examples are described in Section 10. First-order inference is developed in Section 11 for the isotropic case and interpreted geometrically in Section 12. Section 13 describes inference in the weighted case. Hypothesis tests for the one- and two-sample models are given in Section 14, extended to one-way analysis of variance (ANOVA), and illustrated with an example of the association of shape with nutrition. Shape differences are described by an affine component in Section 15. Lastly Section 16 notes some extensions of the models to longitudinal and regression data, and provides a brief overview of the Procrustes approach contrasted with other techniques.

TABLE 1
Notation

N	Number of landmarks
K	Dimension of space
m	Shape dimension, $m = NK - \frac{1}{2}K(K+1) - 1$
L	Number of figures
M	Number of samples
X, Y	Figures
Z	Preshape
$[X]$	Shape of X
X'	X after superimposition
T_X	Procrustes tangent space at X
Σ_K^N	Shape space
$S\Sigma_K^N$	Size-and-shape space
$P\Sigma_K^N$	Preshape space
$R\Sigma_K^N$	Reflection shape space
$SR\Sigma_K^N$	Reflection size-and-shape space
$\boldsymbol{\gamma}$	K-vector for translation
β	Scale factor
Γ	$K \times K$ rotation matrix
$(\boldsymbol{\gamma}, \beta, \Gamma)$	Similarity transformation
$(\boldsymbol{\omega}, \alpha, \Omega)$	Inverse similarity transformation
$[\mu]$	Population mean shape
μ	Figure with shape $[\mu]$
E	Gaussian displacement, e.g. from μ
R	Displacement E after superimposition
P	E in double-rotated co-ordinates
S	R in double-rotated co-ordinates, P after superimposition
G or $G_{\boldsymbol{\gamma}, \beta, \Gamma}$	Generalized Procrustes sum of squares for L figures superimposing by similarity transformations
$G_{\boldsymbol{\gamma}, \Gamma}$, G_Γ	Generalized Procrustes sum of squares superimposing by Euclidean transformations or by rotation only
$G_{\boldsymbol{\gamma}, \beta; \Gamma}$	Generalized Procrustes sum of squares for L figures with standardized location and scale, superimposing using rotation only
G^*	Ordinary Procrustes sum of squares between 2 figures; same usage of subscripts as for G
Σ^{M}	Model metric or covariance (sometimes just Σ)
Σ^{S}	Superimposition metric (sometimes just Σ)

2. PROCRUSTES ESTIMATION OF MEAN SHAPE

We turn now to the geometrical basis of Procrustes methodology, and its relationship to the geometry of shape spaces. Two basic statistical quantities, means and residuals, are defined geometrically for samples of shapes in this section and Section 3. Related distribution theory is deferred to Section 5. The discussion can be readily modified to treat size-and-shape, reflection shape and reflection size-and-shape.

Geometrically, a mean shape is near the centre of the empirical distribution in shape space. For preshapes $Z_1, \ldots, Z_L$, a Procrustes mean shape is $[\bar{Z}]$, where we choose rotations Γ_i and preshape $\bar{Z}$ simultaneously to minimize, by *generalized Procrustes analysis* (GPA) (Gower, 1975),

$$\sum_{i=1}^{L} \| Z_i \Gamma_i - \bar{Z} \|^2. \tag{2.1}$$

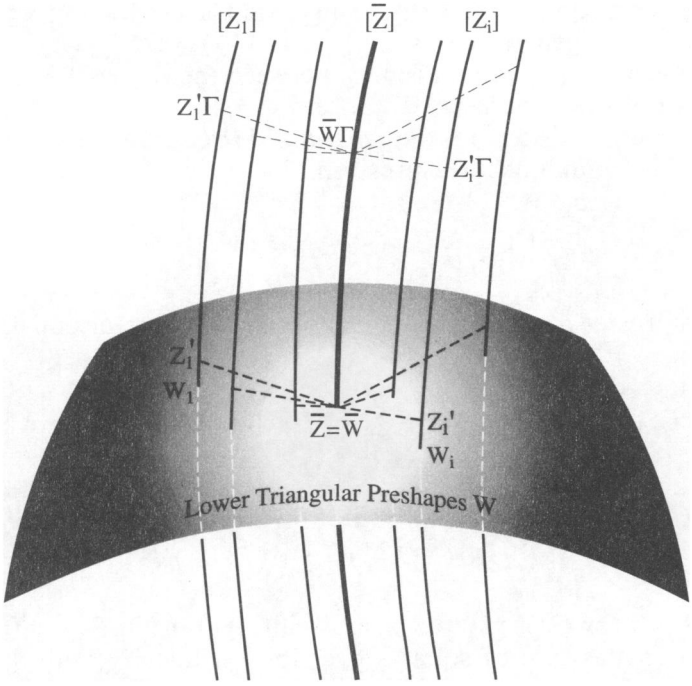

Fig. 2. Schematic diagram of preshape space: the fibres (full lines) of the shapes $[Z_1]$, . . ., $[Z_i]$, . . . are shown intersecting the manifold of lower triangular matrices at preshapes W_i; the mean $\overline{W}$ is the preshape on the manifold with the smallest sum of squared distances to these fibres; the closest preshapes to $\overline{W}$ are at $Z_i' = W_i\Gamma_i$ (joined to $\overline{W}$ by bold broken lines); $\overline{W}$ is on the fibre of the mean shape $[\overline{Z}]$; the entire construction may be rotated by an arbitrary Γ (broken lines)

For uniqueness we may choose $Z_i = W_i$ and $\overline{Z} = \overline{W}$, lower triangular matrices. Fig. 2 shows a section through preshape space that contains the fibres $\{W_i\Gamma_i, \Gamma_i \in SO(k)\}$ and a (continuous) manifold through the lower triangular matrices near $\overline{W}$. We use a prime to denote the closest preshape to $\overline{Z}$ on each fibre, i.e. $Z_i' = Z_i\Gamma_i = Z_i'(Z_i, \overline{Z}, \Gamma)$. ($\overline{Z}$ expresses the dependence of Z_i' both on the Z_i and on the choice of $\overline{Z}$ on its fibre.) $\overline{W}$ minimizes expression (2.1) with $\overline{Z}$ in the manifold; a parallel diagram is obtained when the Z_i' and $\overline{W}$ are each rotated by a fixed rotation Γ.

The minimum of expression (2.1) is the sum of squared Procrustes *distances* between each Z_i and $\overline{Z}$. The Procrustes *sum of squares* between two figures X_1 and X_2 is

$$G^*_{\gamma,\beta,\Gamma}(X_1, X_2) = \min_{\gamma,\beta,\Gamma} \| X_2 - \beta X_1\Gamma - \mathbf{1}_N\gamma^T \|^2. \tag{2.2}$$

The minimum is found by ordinary Procrustes analysis (OPA). The subscript γ, β, Γ indicates that equation (2.2) is the least squares distance between X_2 and the orbit of X_1 under the similarity group. Other subscripts are possible, and a semicolon, where present, is used to separate the components that have been standardized in the decomposition in Section 1 from those that are found by minimization. With these extensions, the squared Procrustes distance (1.6) is

$$d^2([Z_1], [Z_2]) = G^*_\Gamma(Z_1, Z_2) = G^*_{\gamma,\beta;\Gamma}(X_1, X_2). \tag{2.3}$$

We reserve the term squared Procrustes distance for $G^*_{\gamma,\beta;\Gamma}$ and we use Procrustes sum of squares more generally. As $G^*_{\gamma,\beta,\Gamma}(X_1, X_2) \neq G^*_{\gamma,\beta,\Gamma}(X_2, X_1)$ (Section 7), $G^*_{\gamma,\beta,\Gamma}(X_1, X_2)$ is not a (squared) distance between the shapes $[X_1]$ and $[X_2]$, and in particular is not the Procrustes distance.

Without the asterisk, G denotes the *generalized Procrustes sum of squares* between $L \geqslant 2$ figures. The minimum of expression (2.1) is

$$G_\Gamma(Z_1, \ldots, Z_L) = G_{\gamma,\beta;\Gamma}(X_1, \ldots, X_L) = \sum_{i=1}^{L} d^2([Z_i], [\bar{Z}]) = \sum_{i=1}^{L} \|Z_i' - \bar{Z}\|^2. \quad (2.4)$$

A different Procrustes mean shape $[\bar{X}]$ is obtained using the full similarity group, with minimum

$$G_{\gamma,\beta,\Gamma}(X_1, \ldots, X_L) = \sum_{i=1}^{L} G^*_{\gamma,\beta,\Gamma}(X_i, \bar{X}) = \sum_{i=1}^{L} \|X_i' - \bar{X}\|^2, \quad (2.5)$$

where $X_i' = X_i'(X_i, X, \boldsymbol{\gamma}, \beta, \Gamma)$ is the closest figure to X on the orbit of X_i and

$$\sum_i \mathrm{tr}(\tilde{X}_i' \tilde{X}_i'^{\mathrm{T}}) = \sum_i \mathrm{tr}(\tilde{X}_i \tilde{X}_i^{\mathrm{T}})$$

for definiteness. G (or G^*), without a subscript, will mean $G_{\gamma,\beta,\Gamma}$ (or $G^*_{\gamma,\beta,\Gamma}$). The generalized Procrustes sum of squares over the Euclidean group (translations and rotations only) and translation group are written $G_{\gamma,\Gamma}(\mathbf{X})$ and $G_\gamma(\mathbf{X})$ respectively, where $\mathbf{X} = (X_1, \ldots, X_L)$.

The figure centroid minimizes the Procrustes sum of squares irrespective of the scale and orientation of the figure (Section 7), so $G_{\gamma,\beta,\Gamma} = G_{\gamma;\beta,\Gamma}$. The results of shape analysis are affected by whether scale is standardized (orbits in preshape space) or not (orbits in preform space), since $G_{\gamma;\beta,\Gamma} \neq G_{\gamma,\beta;\Gamma}$. It is tempting to remove location and scale once and for all. Thus a considerable literature exists on the definition of scale measures that are stochastically independent of shape (Mosimann, 1970; Sampson and Siegel, 1985; Bookstein and Reyment, 1989) and related questions of allometry (Huxley, 1932). The choice of size statistic (1.3) turns out to be optimal in those terms. Recently attention has turned to alternatives to simply centring the data to remove location (Siegel and Benson, 1982; Small, 1989) (see Section 7). Other choices of objective function are implicit in Mardia and Dryden (1989a, b), Kendall and Le (1990), Kent (1990) and Bookstein (1986, 1991).

Given the optimal Γ_i, and the $\boldsymbol{\gamma}_i$, β_i (fixed or not), the mean shape is seen to be $[\bar{X}]$, where

$$\bar{X} = \frac{1}{L} \sum_{i=1}^{L} X_i'. \quad (2.6)$$

The choices of preshapes $Z_i\Gamma_i$, and more generally of the figures $X_i' = \beta_i X_i \Gamma_i + \mathbf{1}\boldsymbol{\gamma}^{\mathrm{T}}$, say, is a *registration* or *superimposition*. $\bar{X}$ is the co-ordinatewise average of the superimposed figures. Thus, generalizing equation (2.6), we can estimate an average shape by superimposing figures in various ways. Three prominent choices are *edge superimposition*, *optimal superimposition* and *resistant superimposition*.

The most straightforward superimposition is to match a given edge between two landmarks of a planar figure to a common origin and direction. This approach, often

called Roentgenographic cephalometry, is widespread in the biomedical literature, where the edge chosen may be considered biologically 'invariant' (Moss *et al.*, 1987). Bookstein shape co-ordinates (1.7) are a later development and refinement. For example Bookstein and Sampson (1990) assume that **u** has a multivariate Gaussian distribution, compute the mean shape for several samples of figures, find geometric components of the shape change (Section 15) and perform tests of significance for mean shape change relative to the sample covariance.

A second class of criteria involves optimal superimposition. Procrustes methods (Sibson, 1978, 1979; Gower, 1970, 1975, 1984), generalized to allow correlated landmarks and non-spherical error distributions, are the principal topic of this paper.

The third class of superimposition criteria is a compromise between the first two. Criteria for *robust/resistant* superimposition have as goal a high quality match of a fraction (more than 50%) of the landmarks, while highlighting local differences in shape and form. The repeated median technique (Siegel and Benson, 1982) provides robust superimposition of a pair of figures for shape comparison. This may be generalized to a superimposition of more than two figures by iteratively superimposing each figure on the *median* of the figures in their current registration (Rohlf and Slice, 1990).

The three superimposition criteria are illustrated in Fig. 3 for a small data set measured by Descomb comprising eight landmarks in the lateral (medial) section of the neural skulls of 10 rats. (So the edges in Fig. 3 represent the bones of the skull, as is shown in the left-hand half of Fig. 1.) The sample is divided into a group of five controls and five rats with hydrocephaly, a condition associated with excessive fluid retention and a pronounced swelling. Each panel of Fig. 3 shows the five registered outlines (thin lines) and the resulting average (bold broken line) of the hydrocephaly rats. For these data the mean shapes using edge and Procrustes superimpositions are virtually identical. The resistant superimposition result differs, mainly because $\bar{X}$ is the co-ordinatewise *median*.

3. SHAPE DIFFERENCES

Unlike the criterion (2.1) defining an average shape $[\bar{X}]$, there are more subtle issues to consider in discussing shape differences. Two important questions are the following. How should we report the difference between two shapes A and B? When is the shape change from A to B equal to that from C to D? The first question addresses the visualization of shape differences; the second is essential to comparing distributions of shapes (Section 6) and to regression (Section 16). We introduce some basic notions in differential geometry to elucidate the proper approach.

An affine space E possesses *complete parallelism* (Crampin and Pirani, 1986), meaning that, given two points $e_A, e_B \in E$, we may draw a vector equal to $e_B - e_A$ at *any* point of E and remain in E. Therefore, given e_A and e_B it is sufficient to summarize their difference as $e_B - e_A$ and to report only that quantity. Differences $e_D - e_C$ and $e_B - e_A$ may be compared without reference to the original $e_A, \ldots, e_D \in E$. Conversely, if the space is *not* affine then we must report the pair (e_A, e_B) or $(e_A, e_B - e_A)$ and be more careful in comparing two differences.

Figure space and preform space are Euclidean and therefore affine. Preshape space, the sphere, is embedded in an affine space but is not itself affine. (The translation group is not part of the isometries of preshape space.) Jupp (1988) discusses residual

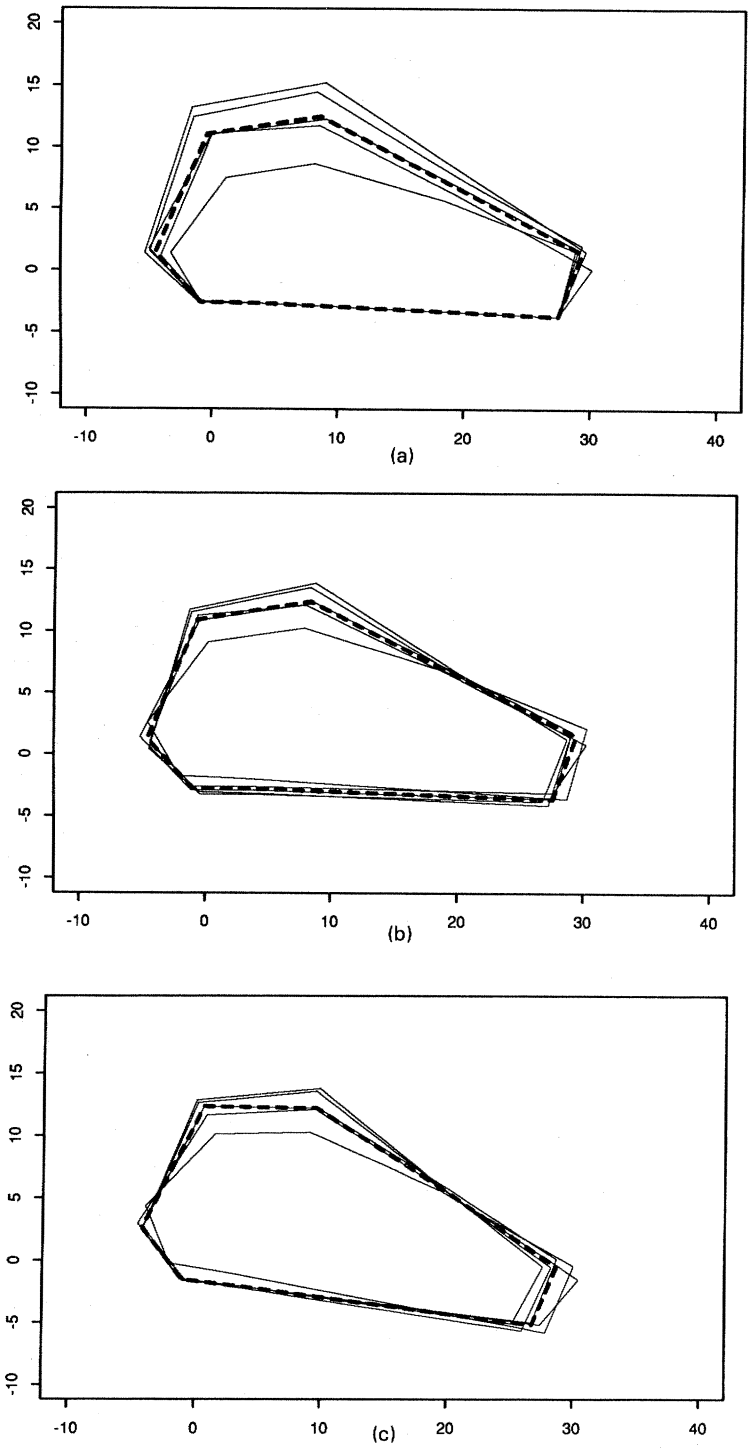

Fig. 3. Three superimpositions of the neural skulls of five rats with hydrocephaly, a localized swelling of the skull: (a) edge superimposition; (b) Procrustes superimposition; (c) resistant superimposition

analysis on the sphere; see also Cox and Snell (1968). Shape space is not readily embedded in Euclidean space in general (exceptions include $K=1$ and $N=3$, $K=2$), but Σ_K^N is differentiable for $K \leqslant 3$, and differentiable almost everywhere otherwise (Kendall and Le, 1990). Thus in comparing shape differences we use local linear approximations to shape space, in terms of either shape co-ordinates (Section 1) *or* the preshape and figure matrices, Z and X.

The stereographic projection of Σ_2^3 to Bookstein shape co-ordinates preserves circles, so that a circularly symmetric shape distribution in one space has circular contours in the other space. For triangles, and planar figures with more landmarks, Bookstein shape co-ordinates are a useful approximation throughout shape space, provided that the largest shape difference in the sample is small relative to the length of $x_2 - x_1$. Bookstein (1984, 1986) summarizes the difference between shapes A and B in a prominent vector $\mathbf{u}_B - \mathbf{u}_A$. He is careful to graph the $N-2$ vectors in $\mathbf{R}^2$ anchored at the respective landmarks of $\mathbf{u}_A$. Such a picture provides a minimal representation of the difference in shape between A and B, and is thus excellent for visualization. The change from A to B is equal to that from C to D when $\mathbf{u}_B - \mathbf{u}_A = \mathbf{u}_D - \mathbf{u}_C$ and $\mathbf{u}_A \approx \mathbf{u}_C$.

The matrices Z and X are elements of an affine space, so that we can report $Z_B - Z_A$ or $X_B - X_A$ (together with Z_A or X_A), provided that the superimposition is appropriate. The minimal requirement that Z or X is a continuous function of shape is met by both Procrustes and edge superimposition. When Procrustes superimposition of X_B or X_A uses the similarity group, the superimposed figure X_B' satisfies the respective rotation and scale constraints derived in Section 12,

$$X_B'^{\mathrm{T}} \tilde{X}_A = \tilde{X}_A^{\mathrm{T}} X_B', \tag{3.1}$$

$$\mathrm{tr}(\tilde{X}_A^{\mathrm{T}} \tilde{X}_B') = \mathrm{tr}(\tilde{X}_B'^{\mathrm{T}} \tilde{X}_B') \tag{3.2}$$

where the tilde denotes centred. Define the *Procrustes tangent space* at the *pole* X_A to be $T_{X_A} = \{X_B' - X_A : X_B \in \mathbf{R}^{NK}\}$. Without the scale constraint (3.2), T_{X_A} is an $(m+1)$-dimensional linear subspace of $\mathbf{R}^{NK}$. With constraint (3.2) T_{X_A} is an m-dimensional approximately linear manifold in $\mathbf{R}^{NK}$, whose geometry approximates the geometry of shape space close to $[X_A]$. The Procrustes sum of squares $G^*(X_C, X_D)$ coincides with the squared Euclidean distance in T_{X_A} when $X_D = X_A$ but not otherwise.

The Riemannian submersion of preshape space to shape space introduces complementary notions of vertical and horizontal in preshape space (Kendall, 1989). A fibre is *vertical* because it maps to a single point. A geodesic in preshape space, i.e. a great circle, is *horizontal* if it has no vertical part; Riemannian distances in preshape space along a horizontal geodesic are equal to Riemannian distances along its image in shape space. Thus, given Z_i, $i=1, \ldots, L$, and the Procrustes mean $\bar{Z}$, the Z_i' lie on a fan of horizontal geodesics about $\bar{Z}$. The Euclidean distances $\|Z_i' - \bar{Z}\|$ in $PT_{\bar{Z}}$, the Procrustes tangent space for preshapes (a subspace of $\mathbf{R}^{(N-1)K}$), equal the corresponding Procrustes distances $d([Z_i], [\bar{Z}])$, but $\|Z_i' - Z_j'\| \neq d([Z_i], [Z_j])$ when Z_i' and Z_j' do not share a horizontal geodesic. Another way of saying this is that the horizontal geodesics in preshape space do not comprise an integral m-dimensional manifold. See Kendall (1984) and Goodall and Mardia (1990b).

How should we compare the shape change A to B with that from C to D, assuming that A is similar to C? One method is to superimpose each figure on a common figure,

the generalized Procrustes mean X_E say, of $X_A - X_D$, and to compare the shape changes in T_{X_E}. An alternative that better preserves the magnitudes of the shape differences is to superimpose X_B on X_A, X_C on X_A and then X_D on X_C', and to compare the shape changes using the embedding of T_{X_A} and $T_{X_C'}$ in $\mathbf{R}^{NK}$. Kendall and Le's (1990) Riemannian metric for shape space may help to compare the methods in detail. A practical application of this methodology is to the two-sample problem (Section 6).

The Procrustes approach is most useful in computing mean shapes and in deciding whether the figures X_A and X_B are random realizations of the same shape. To visualize shape differences a flexible exploratory approach is needed, with goal an informative simultaneous summary of the two shapes. One approach is to express the shape difference as a smooth deformation displayed as a biorthogonal grid (Bookstein, 1991; Bookstein and Sampson, 1990). A different approach is to display the change in position of a small subset of landmarks relative to another subset (perhaps including all the other landmarks) whose shape changes little. Procrustes superimposition does not often reveal such a pattern, but an edge superimposition or resistant technique may be successful. In Fig. 4 we attempt to visualize the difference between the control and hydrocephaly Procrustes means in Descomb's data, using edge, Procrustes and resistant superimposition.

Of the three superimposition criteria, edge superimposition is the most straightforward but robust superimposition may have most potential. The repeated median superimposition readily highlights the localized regions of large shape differences found in biomedical applications (e.g. Fig. 4). Edge superimposition is likewise effective when the edge lies in a region where the shape difference is small. The unweighted Procrustes technique is optimal for Gaussian errors and offers second-order gains in efficiency over other superimpositions (Sections 9 and 16). However, in the weighted least squares and iteratively reweighted least squares versions described later, Procrustes superimposition embraces both edge superimposition and robust estimation.

4. ISOMETRIES OF SHAPE SPACES

Both the *QR* decomposition and the SVD are important in shape theory. The SVD appears in the solution of the Procrustes problem (see Section 7, Seber (1984) and Ten Berge (1977)), the derivation of its statistical properties (see Section 11 and Sibson (1978), in Kendall and Le's (1990) derivation of the Riemannian metrics of shape spaces and in Kendall's (1990) applications of stochastic geometry to shape theory. To help to motivate the models, estimation and inference later, we now use the SVD to describe the isometries of shape spaces.

The SVD of the $N \times K$ centred figure matrix $\tilde{X}$ appears in Section 11. The closely related SVD of the preshape Z is

$$Z = U^R \Lambda V^{RT} = U \Lambda V^T \tag{4.1}$$

where Λ is an $(N-1) \times K$ matrix of zeros apart from the first $d \leqslant \min(N-1, K)$ diagonal elements, $\Lambda_{ii} = \lambda_i > 0$, $i = 1, \ldots, d$, $\lambda_1 \geqslant \lambda_2 \geqslant \ldots \geqslant \lambda_d$. For reflection shapes, $V^R \in O(K)$ and $U^R \in SO(N-1)$. For shapes $V \in SO(K)$ and $U \in O(N-1)$. The singular values λ_i, $i = 1, \ldots, K$, define a point on Kendall and Le's (1990) *shape disc*, which topologically is seen to be the unit sphere S^{K-1} modulo the permutation group and is obtained by a double-Riemannian submersion of each preshape.

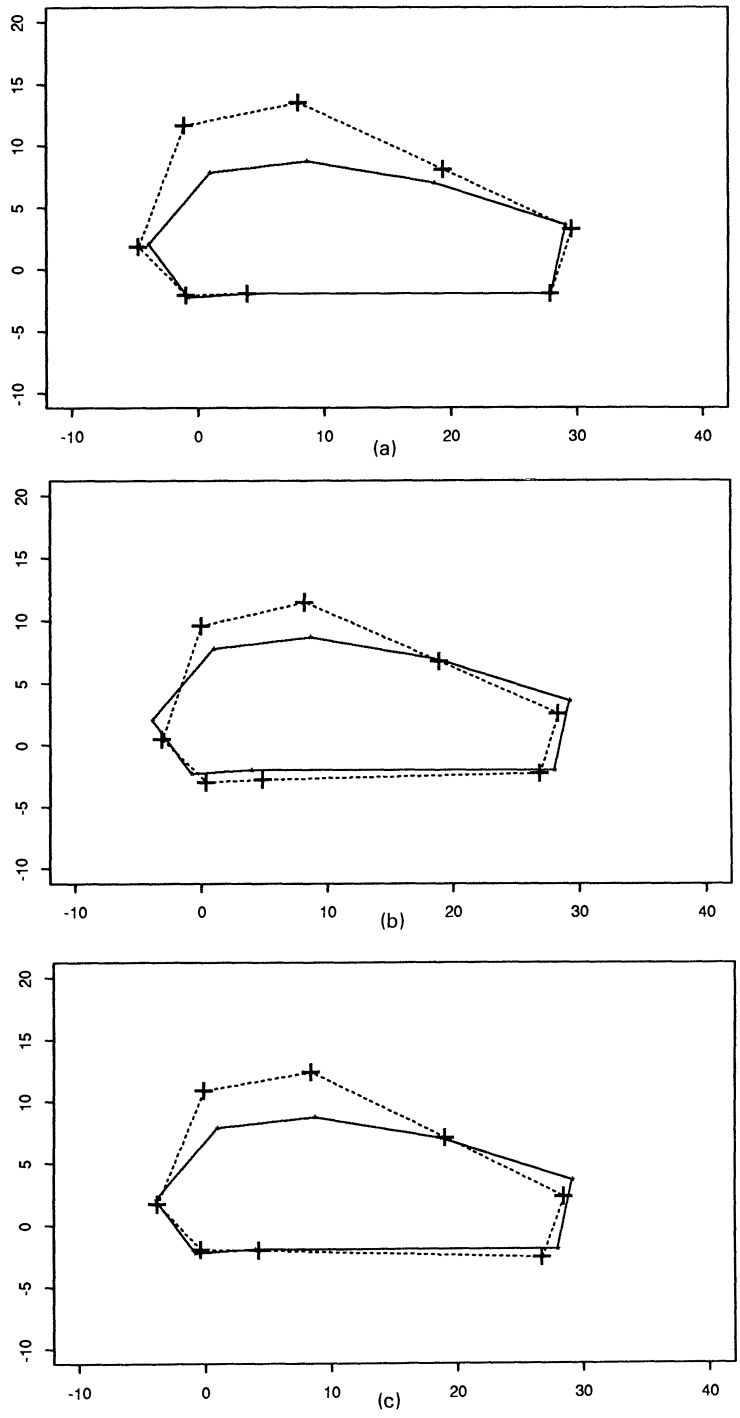

Fig. 4. Three alternative superimpositions of the Procrustes means of the control (——) and hydrocephaly (--------) rats (the edge superimposition is particularly effective for visualization; the Procrustes superimposition is ineffective): (a) edge superimposition; (b) Procrustes superimposition; (c) resistant superimposition

The λ_i, combined with Stiefel manifold co-ordinates from the first K columns of U, comprise another system of shape co-ordinates.

Both the QR decomposition and the SVD separate the shape W and (U, Λ) components of Z from rotation along the fibres of Z by Γ and V^{T}. Additionally the SVD reveals the general automorphism structure of shape space. Premultiplication of each preshape by an orthogonal matrix $U \in O(N-1)$ sends fibres to fibres, horizontal geodesics to horizontal geodesics and is an isometry of preshape space. Thus U induces an automorphism, in fact an isometry, of shape space. For $K \geqslant 3$, $O(N-1)$ is the largest group of isometries of shape space. An arbitrary isometry of preshape space, in $O\{(N-1)K\}$, does not send fibres to fibres and horizontal geodesics to horizontal geodesics. $U \in O(N-1)$ preserves the singular values (leaves the shape disc invariant) so that shape space for $K \geqslant 3$ is inhomogeneous. For example, when $K \geqslant 3$ a planar triangle can be rotated to its reflection. Thus the subspace of planar triangles is the sphere, $S^2(\frac{1}{2})$, with a 2:1 identification of the north and south hemispheres, and the equator of collinear triangles is then an 'edge'. Incidentally this example helps to characterize reflection shape space for planar triangles.

When $K=2$ the automorphism group includes any rotation in $O(2N-2)$ that multiplies the rows of Z by *different* planar rotations. These 2×2 rotations *commute* with Γ and V^{T}, so the resulting transformation preserves fibres and horizontal geodesics. Combining the two groups we are able to map any shape in Σ_2^N to any other by an isometry, and $\mathbf{C}P^{N-2}(4)$ is said to be homogeneous, as noted by Kendall (1984).

In summary, the isometries of shape space for $K \geqslant 3$ are induced by preshape rotations given by the Kronecker products $U^* \otimes V^*$

$$U^* \otimes V^*\colon\ Z \rightarrow U^* Z V^{*\mathrm{T}} \tag{4.2}$$

where $U^* \in O(N-1)$ and $V^* \in SO(K)$. Multiplication on the right-hand side by V^* is simply rotation along a fibre, and $U^* \otimes V^*$ is the $(N-1)K \times (N-1)K$ linear operator $(U^* \otimes I)(I \otimes V^*) = (I \otimes V^*)(U^* \otimes I)$ that multiplies $\mathrm{vec}(Z)$ from the left. For $K=2$ the isometries are induced by a *generalized* Kronecker product. One definition is

$$U^* \otimes \{V^*\}_{N-1}\colon\ \mathbf{z}_i \rightarrow V_i^* \mathbf{z}_i' \tag{4.3}$$

where $U^* \in O(N-1)$, $\mathbf{z}_i$ is the ith row of Z, $\mathbf{z}_i'$ is the ith row of U^*Z and $\{V^*\}_{N-1} = \{V_1^*, \ldots, V_{N-1}^*\}$ is a set of 2×2 rotation matrices. $U^* \otimes \{V^*\}_{N-1}$ is the linear operator $\oplus V_i^* (U^* \otimes I)$ where $\oplus V_i^*$, the direct sum of the V_i^*, is a block diagonal matrix. The direct sum does not commute with $U^* \otimes I$, and an alternative to definition (4.3) is $(U^* \otimes I) \oplus V_i^*$; see also Regalia and Mitra (1989).

We conclude with an interesting connection between the QR decomposition, the SVD and Procrustes superimposition when $K=2$: transform preshape space using definition (4.3) so that, for given Z_{A}, $(Z_{\mathrm{A}})_{11} = 1$ and $(Z_{\mathrm{A}})_{ij} = 0$ otherwise. Then, for any other preshape Z_{B}, W_{B} is the Procrustes superimposition on Z_{A}, because $(W_{\mathrm{B}}\Gamma)_{11}$ is a maximum when $\Gamma = I_K$ (see Fig. 2).

5. GAUSSIAN MODELS FOR LANDMARK DATA

Let $X_1, \ldots, X_L$ contain the co-ordinates of N landmarks in $\mathbf{R}^K$. In a Gaussian model for landmark data, each figure X_i is modelled as a similarity transformation

$(\boldsymbol{\omega}_i, \alpha_i, \Omega_i)$ of a zero-mean Gaussian displacement $\text{vec}(E_i) \sim N(0, \Sigma)$ of the unknown population mean μ,

$$X_i = \alpha_i(\mu + E_i)\Omega_i + \mathbf{1}_N \boldsymbol{\omega}_i^{\mathrm{T}}. \tag{5.1}$$

Equivalently, let $(\boldsymbol{\gamma}_i, \beta_i, \Gamma_i)$ be the inverse transformation, $(\boldsymbol{\omega}_i, \alpha_i, \Omega_i) = (-\beta_i^{-1}\Gamma_i\boldsymbol{\gamma}_i, \beta_i^{-1}, \Gamma_i^{\mathrm{T}})$. Then

$$X_i' = \beta_i X_i \Gamma_i + \mathbf{1}_N \boldsymbol{\gamma}_i^{\mathrm{T}} \sim N(\mu, \Sigma). \tag{5.2}$$

Model (5.1) can be restated: given μ and E_i, we observe $[\mu + E_i]$. The order of model (5.1) (displace μ, then transform) ensures that the displacements have common orientation and scale relative to μ. The underlying picture is of a geometrical figure whose landmarks are displaced L times and each result is measured in a different unknown co-ordinate system specified by the equivalent sets of parameters $(\boldsymbol{\omega}_i, \alpha_i, \Omega_i)$ and $(\boldsymbol{\gamma}_i, \beta_i, \Gamma_i)$. The *population mean shape* $[\mu]$ is well defined but μ is defined only to within a similarity transformation. (The likelihood has ridges.) The actual choice of μ emerges during estimation, testing and visualization. Thus model (5.1) is a model for shapes, not for figures. With appropriate modifications ($\Omega_i \in O(K)$, no α_i scale components) model (5.1) becomes a model for reflection shapes, forms or reflection forms.

We assume that the E_i: $N \times K$ are independent. A common simplifying assumption is $\Sigma = I_{NK}$, the identity matrix, in which case the displacements are independent and identically distributed (IID) spherical Gaussian. Much of the inference for Procrustes methods (Section 11) is first order in ϵ, where $\text{var}(E_i) = \epsilon^2 I_{NK}$. Thus Sibson (1978, 1979) uses the slightly more general model in which errors are Gaussian to first order.

Consistent with Section 3, each displacement E_i is defined relative to the population mean μ. The covariance Σ is defined relative to μ also. From a modelling perspective the covariance, or *model metric*, accommodates two sources of variation in the data. These are

(a) measurement error in digitizing the landmark co-ordinates and
(b) intrapopulation variation in shape.

The E_i in model (5.1) are the proper model for measurement error. We assume that the intrapopulation variability, or shape density, can be expressed by such displacements also, at least to first order. A random displacement E_i of μ, however, includes components for both shape differences and similarity transformations: between fibres (in T_μ) and along fibres. If μ and X_i are on the same fibre then the displacement has only a similarity transformation component. Typically μ and X_i lie on different fibres, and a component of each E_i is confounded with the $(\boldsymbol{\gamma}_i, \beta_i, \Gamma_i)$. (Anticipating Section 10, when Σ is estimated from the residuals *after* superimposition, the components along fibres are zero to first order.)

The prototypical estimation problem is to estimate $[\mu]$, given $X_1, \ldots, X_L$. Two alternative treatments of the unknown parameters $(\boldsymbol{\gamma}_i, \beta_i, \Gamma_i)$ lead to a marginal and a full MLE of $[\mu]$. In marginal estimation the explicit $(\boldsymbol{\gamma}_i, \beta_i, \Gamma_i)$ can be omitted from model (5.2), so the usual set-up states that each landmark $\mathbf{x}_i$ has an

independent Gaussian distribution with mean $\boldsymbol{\mu}_i$ and covariance Σ_i. The *shape density*, which is the integral of the Gaussian density in figure space over the orbit of the similarity group, has been computed extensively for planar figures.

With the μ_i equal and isotropic errors, $\Sigma_i = \epsilon^2 I$, Kendall (1984) shows that the shape density is uniform. Small (1981, 1984, 1988) and Kendall (1984) discuss the IID case more generally. Bookstein (1984) argues that when the $\boldsymbol{\mu}_i$ are unequal and $\Sigma_i = \epsilon^2 I$, then, for triangles $\mathbf{u}$ is approximately normal. The exact shape density for triangles is given by Mardia (1989a) and for arbitrary N by Mardia and Dryden (1989a, b). Mardia (1989a) provides a more general discussion for planar triangles, while Dryden (1989) and Dryden and Mardia (1991) compute the shape distributions for unequal Σ_i. A stochastic calculus approach has been taken by Le (1990b) and a geometrical approach by Goodall and Mardia (1990b), who integrate out location, scale and rotation in stages in Kendall's spaces. Goodall and Mardia (1990a) extend shape distributions to $K > 2$ in special cases, through the connections between non-central multivariate statistical theory and shape theory: integration over the orthogonal group (former) coincides with integration over the special orthogonal group (latter) when rank $\mu < K$. In particular Bartlett's generalized decomposition gives the shape density immediately whenever rank $\mu = 1$, which when $K = 2$ includes all μ because of the homogeneity of Σ_2^N.

Full likelihood estimates, with the $(\boldsymbol{\gamma}_i, \beta_i, \Gamma_i)$ as nuisance parameters, are computed by Procrustes methods. The two estimates differ for moderately dispersed shapes, a fact verified empirically by Ian Dryden and myself. The contribution of a figure X to the joint likelihood is either the integral of the figure space density over the orbit of the similarity group (marginal) or the maximum over that orbit (full): see Fig. 2. For translations only and Gaussian errors, the distinction is vacuous; with the inclusion of scale and rotation the difference is sizable when the shapes of the X_i are widely dispersed.

Which of the two approaches is the more realistic? In terms of measurement error both alternatives are perhaps too extreme: data come with a perturbation to the co-ordinate frame but also with clues, at least partially independent of the landmark data, to the correct registration. An example of such clues is the outline of bones, between the landmarks which are at sutures of the bones: Fig. 1. (The correct registration is a construct of the model and is not estimable; it is $\mu + E_i$. Unless the shapes of the X_i are identical there is no common co-ordinate system apart from this construct.) A more realistic model may have a Bayesian flavour, involving an explicit non-diffuse prior on the similarity transformations, for example based on a technician's use of informal edge superimposition, or some other visually estimated procedure, in selecting the co-ordinate system for each figure.

We do not have extensive data to use to choose between marginal and full MLEs given variation in shape (b). Further alternatives are the similarity invariant complex Bingham distribution for shapes of planar configurations (Kent, 1990) and the template models of Chow *et al.* (1988) involving probability measures on polygons.

5.1. *Choice of* Σ

Motivated by the isometries of shape spaces, we emphasize factored covariances Σ of the form

$$\Sigma = \Sigma_N \otimes \Sigma_K \tag{5.3}$$

where Σ_N: $N\times N$ and Σ_K: $K\times K$. If $\Sigma_N^{-1}=Q^{\mathrm{T}}Q$, the Cholesky decomposition say, then each row of QE_i is IID with covariance Σ_K. Thus Σ_N and Σ_K model the covariance between landmarks and non-spherical displacement at each *derived landmark* respectively. When $K=2$ there are several generalizations to consider, the most attractive being when the rows of QE_i are independent with covariance Σ_{Kj}, $j=1, \ldots, N$, for some choice of Q. Then

$$\Sigma=(Q^{-1}\otimes I)\bigoplus_{j=1}^{N}\Sigma_{Kj}(Q^{-\mathrm{T}}\otimes I). \tag{5.4}$$

For identifiability we require, for $K\geqslant 3$, $\mathrm{tr}(\Sigma_N)=1$ and, for $K=2$, $\mathrm{diag}(\Sigma_N)=I_N$ (so Σ_N is a correlation matrix). The factored structure (5.3) reduces the number of covariance parameters from $NK(NK+1)/2$ to $K(K+1)/2+N(N+1)/2-1$. There are $N(N-1)/2+NK(K+1)/2$ covariance parameters in model (5.4). Table 2 gives counts of the covariance and model parameters for several choices of N and K. In practice, in estimating Σ from the data, we must restrict models (5.3) and (5.4) to a simple structure, such as $\Sigma_N=I$ or Σ_N diagonal, unless L is exceptionally large.

For model (5.3) with $\beta_i=1$ and $\Gamma_i=I$, the factored covariance between the kth and mth co-ordinates of respectively the jth and lth landmarks of X_i is

$$\mathrm{cov}\{(X_i)_{jk},\ (X_i)_{lm}\}=(\Sigma_N)_{jl}(\Sigma_K)_{km}. \tag{5.5}$$

Thus landmarks tend to perturb in parallel directions. As a model for measurement error, equation (5.5) is a drawback, as the direction of greatest variation may differ markedly between landmarks. Fig. 5 shows data of 10 undernourished rats (details are given in Section 14), superimposed on the Procrustes mean. The variation at the landmarks tends to be both parallel and perpendicular to the sutures between bones. This pattern is accommodated by model (5.4). As a model for shape (b), the parallel displacements in model (5.5) amount to a *shearing* component, which are part of the natural affine family of deformations of a figure (Section 15).

Procrustes methods yield MLEs (Section 10) when the covariance has the form (5.3) or (5.4). When $K=2$ and the covariance is completely general then a weighted multiple-regression technique may be used. We now summarize Procrustes analysis of a one-sample model and develop a two-sample model. The two-sample model is easily extended to one-way ANOVA in Section 14.

TABLE 2
Covariance degrees of freedom for selected N, K and Σ

N	K	NK	m	$\Sigma_N\otimes I_K$	$I_N\otimes\Sigma_K$	$\oplus\Sigma_{Kj}$	*Equation (5.3)*	*Equation (5.4)*	Σ
3	2	6	2	6	3	9	8	12	21
8	2	16	12	36	3	24	38	52	136
20	2	40	36	210	3	60	212	250	820
40	2	80	76	820	3	120	822	900	3240
4	3	12	5	10	6	24	15	30	78
10	3	30	23	55	6	60	60	105	465

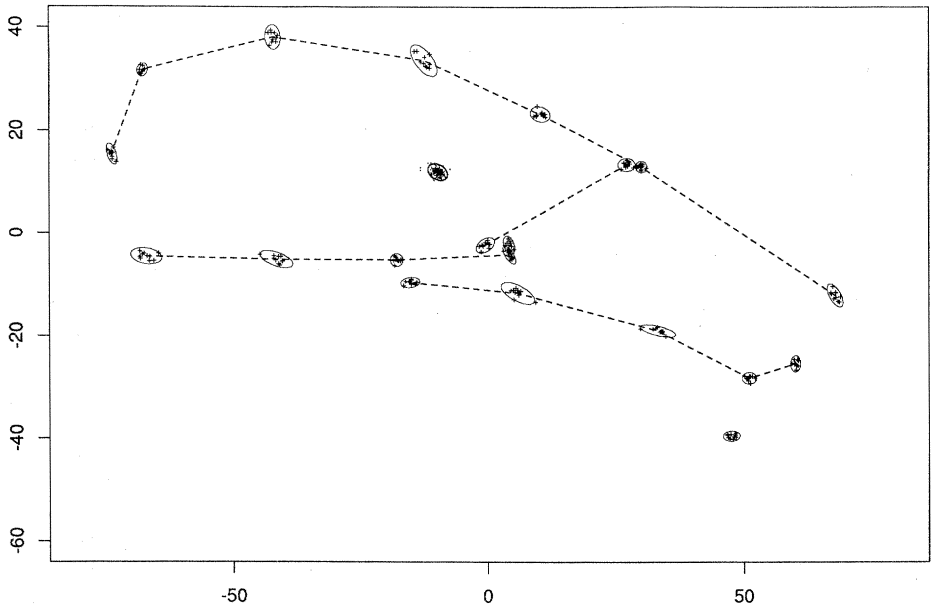

Fig. 5. Pucciarelli data of 10 close-bred undernourished rat pups at 21 days superimposed by GPA: the 10×20 superimposed landmarks are shown, together with ellipses containing 95% of a fitted bivariate Gaussian distribution at each landmark, and for all landmarks together (inside)

6. PROCRUSTES ANALYSIS IN ONE- AND TWO-SAMPLE MODELS

6.1. *One-sample Model*

Given $X_1, \ldots, X_L$ and model (5.1), we may wish to compare the population mean shape $[\mu]$ with the shape $[\nu]$ of the fixed figure ν say (e.g. an equilateral triangle when $N=3$). We first obtain an estimate $[\hat{\mu}]$ of $[\mu]$ by superimposition/maximum likelihood. The comparison may include visualization of the difference in shape between $[\hat{\mu}]$ and $[\nu]$, a formal test (Section 14) and, if there is a significant shape difference, some parametric descriptions and testing of it (Section 15). We may superimpose ν on $\hat{\mu}$ (or $\hat{\mu}$ on ν), transforming ν to ν' say. One or more superimpositions may help to visualize the shape difference, but a specific, optimal, superimposition is needed to test H_0: $[\mu] = [\nu]$ *versus* H: $[\mu] \neq [\nu]$. For this we choose ν' to make the residual $R = \nu' - \hat{\mu}$ as small as possible in the model metric Σ. We minimize $\|R\|_\Sigma = \mathbf{r}^{\mathrm{T}} \Sigma^{-1} \mathbf{r}$ where $\mathbf{r} = \mathrm{vec}(R)$. The formal test of significance is an F-ratio test or Hotelling's T^2-test in the tangent space $T_{\hat{\mu}}$.

6.2. *Two-sample Model: Cross-sectional Data*

Let $X_1, \ldots, X_{L_X}$ and $Y_1, \ldots, Y_{L_Y}$ be $N \times K$ co-ordinate matrices of samples of L_X and L_Y figures from two populations. (We use the subscripts X and Y where necessary.) Model (5.1) becomes

$$X_i = \alpha_{Xi}(\mu_X + E_{Xi})\Omega_{Xi} + \mathbf{1}_N \boldsymbol{\omega}_{Xi}^{\mathrm{T}} \tag{6.1}$$

$$Y_i = \alpha_{Yi}(\mu_Y + E_{Yi})\Omega_{Yi} + \mathbf{1}_N \boldsymbol{\omega}_{Yi}^{\mathrm{T}}. \tag{6.2}$$

We assume that $\text{cov}\{\text{vec}(E_{Xi})\} = \Sigma_X$, $\text{cov}\{\text{vec}(E_{Yi})\} = \Sigma_Y$ and that the E_{Xi} and E_{Yi} are independent. In comparing the two populations we shall superimpose one mean estimate on the other, in this case $\hat{\mu}_X$ on $\hat{\mu}_Y$ say, with one or more superimpositions to help to visualize the shape differences and another, optimal, superimposition to test the hypothesis H_0: $[\mu_X] = [\mu_Y]$ *versus* H: $[\mu_X] \neq [\mu_Y]$. The optimal superimposition makes $\hat{\mu}_X' - \hat{\mu}_Y$ as small as possible relative to both sets of errors.

Suppose that $\hat{\mu}_X'$ is the similarity transformation $(\boldsymbol{\gamma}, \beta, \Gamma)$ of $\hat{\mu}_X$, and likewise define $\mu_X' = \beta\mu_X\Gamma + \mathbf{1}_N\boldsymbol{\gamma}^T$. Then model (6.2) becomes

$$X_i = \alpha_{Xi}'(\mu_X' + E_{Xi}')\Omega_{Xi}' + \mathbf{1}_N\boldsymbol{\omega}_{Xi}'^T \tag{6.3}$$

where

$$(\boldsymbol{\omega}_{Xi}', \alpha_{Xi}', \Omega_{Xi}') = (\boldsymbol{\omega}_{Xi} - \alpha_{Xi}\Omega_{Xi}^T\boldsymbol{\gamma}/\beta, \alpha_{Xi}/\beta, \Gamma^T\Omega_{Xi}),$$

the composition of $(\boldsymbol{\omega}_{Xi}, \alpha_{Xi}, \Omega_{Xi})$ and $(\boldsymbol{\gamma}, \beta, \Gamma)$. The displacement after superimposition is $E_{Xi}' = \beta E_{Xi}\Gamma^T$, and the transformed covariance is

$$\begin{aligned}\Sigma_{X'} &= \text{cov}\{\text{vec}(E_{Xi})\} \\ &= \beta^2(I \otimes \Gamma)\Sigma_X(I \otimes \Gamma^T) \\ &= \beta^2\Sigma_{XN} \otimes \Gamma\Sigma_{XK}\Gamma^T,\end{aligned} \tag{6.4}$$

say. $E_{Xi} \sim N(0, \epsilon_X^2 I)$ implies that $E_{Xi}' \sim N(0, \beta^2\epsilon_X^2 I)$ since Γ is orthogonal.

The distribution of $R = \hat{\mu}_X' - \hat{\mu}_Y$ depends on $\Sigma_{X'}$ and Σ_Y. The formal test of significance (Section 14) is an F-ratio test or Hotelling's T^2-test in the union of the subspaces $T_{\hat{\mu}_{X'}}$ and $T_{\hat{\mu}_Y}$. For simplicity we may assume that $\Sigma_{X'}$ and Σ_Y are equal; for example when the errors are isotropic this implies that $\epsilon_Y = \beta\epsilon_X$. Methods for unequal Σ appear in Anderson (1984), section 5.5. In any case, the two tangent spaces $T_{\hat{\mu}_X}$ and $T_{\hat{\mu}_Y}$ must almost coincide.

7. PROCRUSTES METHODS FOR ESTIMATION

In this and the following section we review the use of Procrustes analysis to obtain weighted least squares (WLS) and iteratively reweighted least squares (IRLS) estimates of the population mean and covariance matrices. (We use the term weighted, instead of generalized, to avoid confusion with GPA, without implying a restriction to diagonal covariances.) When each covariance matrix is the identity, unweighted OPA and GPA are used. We discuss WLS estimation for the models (5.3) $\Sigma = \Sigma_N \otimes \Sigma_K$ and its generalization (5.4) when $K = 2$. WLS for $\Sigma_N \neq I$ is straightforward. The algorithm when $\Sigma_K \neq I$ is more difficult, as the covariance and orthogonal mapping are mismatched. Following the discussion of weighted Procrustes methods, we note that the Procrustes Σ *need not* match the model Σ (Section 9) and describe sequences of steps for fitting the one- and two-sample models when the covariances are known (WLS) or unknown (IRLS) in Section 10.

7.1. *Ordinary Procrustes Analysis*

Let X and Y be two figures containing N landmarks in $\mathbf{R}^K$. Parallel to models (5.1) and (5.2), write

$$X = \alpha(Y + E)\Omega + \mathbf{1}_N\boldsymbol{\omega}^T, \tag{7.1}$$

$$Y = \beta X\Gamma + \mathbf{1}_N\boldsymbol{\gamma}^T - E \tag{7.2}$$

where $(\boldsymbol{\omega}, \alpha, \Omega)$ and $(\boldsymbol{\gamma}, \beta, \Gamma)$ are inverses. Model (7.2) is the multivariate linear regression, or affine, model with the restriction that $\beta\Gamma$ is a multiple of an orthogonal matrix. Let $\mathbf{e} = \text{vec}(E)$. The weighted Procrustes, WLS, estimates $(\hat{\beta}, \hat{\Gamma}, \hat{\boldsymbol{\gamma}})$ minimize the L_2-norm $\| Y - \beta X\Gamma - \mathbf{1}_N\boldsymbol{\gamma}^{\mathrm{T}} \|_{\Sigma}^2 = \mathbf{e}^{\mathrm{T}}\Sigma^{-1}\mathbf{e}$. With $X' = \hat{\beta}X\hat{\Gamma} + \mathbf{1}_N\hat{\boldsymbol{\gamma}}^{\mathrm{T}}$ and $R = X' - Y$, the weighted Procrustes sum of squares is written $G_S^*(X, Y; \Sigma) = \mathbf{r}^{\mathrm{T}}\Sigma^{-1}\mathbf{r}$.

For any K and factored covariance model (5.3), we minimize

$$\text{tr}(Y - \beta X\Gamma - \mathbf{1}_N\boldsymbol{\gamma}^{\mathrm{T}})^{\mathrm{T}}\Sigma_N^{-1}(Y - \beta X\Gamma - \mathbf{1}_N\boldsymbol{\gamma}^{\mathrm{T}})\Sigma_K^{-1} \tag{7.3}$$

by OPA, written OPA(Σ). The unweighted Procrustes technique, when $\Sigma_N = I_N$, $\Sigma_K = I_K$, can be traced to Mosier (1939). Sibson (1978, 1979) considers OPA in detail. Mardia *et al.* (1979), Gower (1984) and Seber (1984) include excellent introductory expositions of OPA. For consistency with the conventional regression model we have interchanged the roles of X and Y relative to these accounts.

In outline, the OPA translation $\hat{\boldsymbol{\gamma}}$ is the difference

$$(Y - \tilde{Y}) - \hat{\beta}(X - \tilde{X})\hat{\Gamma} = \mathbf{1}_N^{\mathrm{T}}\hat{\boldsymbol{\gamma}} \tag{7.4}$$

where $\tilde{X}$ and $\tilde{Y}$ are centred. We now choose $\hat{\beta}$ and $\hat{\Gamma}$ to minimize $\| \tilde{Y} - \beta\tilde{X}\Gamma \|^2$. Expanding this expression,

$$\| \tilde{Y} - \beta\tilde{X}\Gamma \|^2 = \text{tr}(\tilde{Y}^{\mathrm{T}}\tilde{Y}) + \beta^2\,\text{tr}(\tilde{X}^{\mathrm{T}}\tilde{X}) - 2\beta\,\text{tr}(\tilde{Y}^{\mathrm{T}}\tilde{X}\Gamma). \tag{7.5}$$

Let the SVD of $\tilde{Y}^{\mathrm{T}}\tilde{X}$ be $L\Delta M^{\mathrm{T}}$. Then for any β the minimizing Γ is

$$\hat{\Gamma} = ML^{\mathrm{T}} = (\tilde{X}^{\mathrm{T}}\tilde{Y}\,\tilde{Y}^{\mathrm{T}}\tilde{X})^{-1/2}\tilde{X}^{\mathrm{T}}\tilde{Y} \tag{7.6}$$

and, given $\Gamma = \hat{\Gamma}$, regression through the origin gives

$$\hat{\beta} = \frac{\text{tr}(\tilde{Y}^{\mathrm{T}}\tilde{X}\hat{\Gamma})}{\text{tr}(\hat{\Gamma}^{\mathrm{T}}\tilde{X}^{\mathrm{T}}\tilde{X}\hat{\Gamma})} = \frac{\text{tr}(\tilde{X}^{\mathrm{T}}\tilde{Y}\,\tilde{Y}^{\mathrm{T}}\tilde{X})^{1/2}}{\text{tr}(\tilde{X}^{\mathrm{T}}\tilde{X})}. \tag{7.7}$$

The estimate $\hat{\Gamma}$ does not depend on the scale of the two figures, but $\hat{\beta}$ depends on Γ. Therefore least squares estimation of the similarity transformation $(\boldsymbol{\gamma}, \beta, \Gamma)$ has the natural order $\boldsymbol{\gamma}$ (mean centring), then Γ, and finally β. On interchanging the roles of X and Y, the estimated rotation is $\hat{\Gamma}^{\mathrm{T}}$ but the scale is *not* $1/\hat{\beta}$. As we have noted, these two Procrustes sums of squares are unequal, $G^*_{\boldsymbol{\gamma},\beta,\Gamma}(X, Y) \neq G^*_{\boldsymbol{\gamma},\beta,\Gamma}(Y, X)$, and also do not coincide with the squared Procrustes distance in Kendall's shape space, $G^*_{\boldsymbol{\gamma},\beta;\Gamma}(X, Y) = G^*_{\boldsymbol{\gamma},\beta;\Gamma}(Y, X)$. This is not surprising, as the three analogous quantities differ in the multivariate regression setting, where they correspond to two different projections and a squared Euclidean distance respectively. What is notable is that, for orthogonal fitting without scale, the three quantities coincide for size-and-shapes, and for spherical data. Model (5.1) shows that the correct approach is asymmetric in X_i and $\bar{X}$. However, some symmetric choices of scale and measures of congruence are those of Sibson (1978) and Robert and Escoufier (1976). Simple linear regression is an interesting special case. When $K = 1$ each of the figures is a vector in $\mathbf{R}^N$, the orthogonal transformation is $\Gamma = \pm 1$ and the Procrustes model for two figures reduces to the simple linear regression model with intercept $\boldsymbol{\gamma}$ and slope $\beta\Gamma$. Two vectors X and Y are affine equivalent if their correlation $R = 1$, and Kendall's shape space Σ_1^N is the sphere $S^{(N-2)}$. When $\| \tilde{X} \| = \| \tilde{Y} \| = 1$ the Procrustes sums of squares are each $1 - R^2$ and the squared Procrustes distance is twice that.

The OPA described above is for reflection shapes, as Kendall (1984) points out. The expression $\text{tr}(\tilde{Y}^{\mathrm{T}}\tilde{X}\Gamma)$ is bounded above by $\text{tr}(\Delta)$ when L and M are orthogonal.

Thus when L and M are special orthogonal the upper bound is $\mathrm{tr}(\Delta)-2\Delta_{KK}$ ($\Delta_{11}\geqslant\Delta_{22}\geqslant\ldots\geqslant\Delta_{KK}$). In practice, when just one of L and M has determinant -1 we multiply the sign of the Kth column of that matrix by -1 before constructing $\hat{\Gamma}$.

The centred co-ordinate matrices $\tilde{X}$ and $\tilde{Y}$ are the components of X and Y orthogonal to the projection by $\mathbf{1}_N\mathbf{1}_N^{\mathrm{T}}/N$. When $\Sigma=\Sigma_N\otimes\Sigma_K$ the weighted least squares estimate of $\boldsymbol{\gamma}$ is determined by the projection

$$B=\mathbf{1}_N(\mathbf{1}_N^{\mathrm{T}}\Sigma_N^{-1}\mathbf{1}_N)^{-1}\mathbf{1}_N^{\mathrm{T}}\Sigma_N^{-1}, \tag{7.8}$$

$$\mathbf{1}_N^{\mathrm{T}}\hat{\boldsymbol{\gamma}}=B(Y-\hat{\beta}X\hat{\Gamma}). \tag{7.9}$$

Equation (7.9) generalizes equation (7.4). The projection minimizes the weighted sum of squares for each column, so minimizes expression (7.3) for all matrices Σ_K. Furthermore, the translation can be eliminated once and for all by projecting each of X and Y separately, as this minimizes expression (7.3) for all choices of $\hat{\beta}$ and $\hat{\Gamma}$:

$$\begin{aligned}\tilde{X}&=(I-B)X,\\ \tilde{Y}&=(I-B)Y.\end{aligned} \tag{7.10}$$

For general Σ the optimal Γ does not depend on β. When $\Sigma_K=I_K$ and Σ_N is non-singular, expression (7.3) becomes

$$\|\tilde{Y}-\beta\tilde{X}\Gamma\|^2_{\Sigma_N\otimes I}=\mathrm{tr}(\tilde{Y}^{\mathrm{T}}\Sigma_N^{-1}\tilde{Y})+\beta^2\,\mathrm{tr}(\tilde{X}^{\mathrm{T}}\Sigma_N^{-1}\tilde{X})-2\beta\,\mathrm{tr}(\tilde{X}^{\mathrm{T}}\Sigma_N^{-1}\tilde{Y}\Gamma). \tag{7.11}$$

We replace X by QX and Y by QY, where $\Sigma_N^{-1}=Q^{\mathrm{T}}Q$ is the Cholesky decomposition, and compute $\hat{\Gamma}$ as before. When $\Sigma_K\neq I$ there is no explicit expression for $\tilde{X}\hat{\Gamma}$, as there is in equation (7.7), the minimum is not unique and the iterative algorithm of Koschat and Swayne (1989) may be used. Given the rotation $\hat{\Gamma}$, weighted regression through the origin yields

$$\hat{\beta}=\frac{\mathrm{vec}(\tilde{X}\hat{\Gamma})^{\mathrm{T}}\Sigma^{-1}\mathrm{vec}(\tilde{Y})}{\mathrm{vec}(\tilde{X}\hat{\Gamma})^{\mathrm{T}}\Sigma^{-1}\mathrm{vec}(\tilde{X}\hat{\Gamma})}=\frac{\mathrm{tr}(\tilde{X}\hat{\Gamma})^{\mathrm{T}}\Sigma_N^{-1}(\tilde{Y})\Sigma_K^{-1}}{\mathrm{tr}(\tilde{X}\hat{\Gamma})^{\mathrm{T}}\Sigma_N^{-1}(\tilde{X}\hat{\Gamma})\Sigma_K^{-1}}. \tag{7.12}$$

When $K=2$ and Σ has the form (5.4), we minimize a generalization of expression (7.3). After premultiplication by Q we may assume that $\Sigma_N=I$. The expression for the weighted least squares estimate of location is relatively complicated and is not reported here. For rotation, write

$$\Sigma_{Kj}^{-1}=D_j^{\mathrm{T}}\begin{pmatrix}w_{1j}^2 & 0\\ 0 & w_{2j}^2\end{pmatrix}D_j \tag{7.13}$$

and $\mathbf{x}_j$, $\mathbf{y}_j$ for the jth row of the centred $\tilde{X}$ and $\tilde{Y}$ respectively. We choose Γ to minimize

$$\Sigma_{j=1}^N(\mathbf{y}_j-\Gamma^{\mathrm{T}}\mathbf{x}_j)^{\mathrm{T}}D_j^{\mathrm{T}}\begin{pmatrix}w_{1j}^2 & 0\\ 0 & w_{2j}^2\end{pmatrix}D_j(\mathbf{y}_j-\Gamma^{\mathrm{T}}\mathbf{x}_j). \tag{7.14}$$

The rotations D_j commute with Γ so that, after transforming the $\mathbf{x}_j$ and $\mathbf{y}_j$ separately, we may assume that each D_j is the identity. Let ψ be the rotation angle of Γ and $t=\cos\psi$. Then omitting the subscripts j and summations, and writing (x_1, x_2) and (y_1, y_2) as landmark co-ordinates, expression (7.14) has extrema at the solutions of

$$\pm\{t(x_2^2-x_1^2)(w_1^2-w_2^2)+x_1y_1w_1^2+x_2y_2w_2^2\}\sqrt{(1-t^2)}$$
$$=(2t^2-1)x_1x_2(w_1^2-w_2^2)-t(x_2y_1w_1^2-x_1y_2w_1^2). \tag{7.15}$$

This generalizes Koschat and Swayne's (1989) result when $K=2$ to the non-IID case.

For general Σ and $K=2$, we can estimate $(\boldsymbol{\gamma}, \beta, \Gamma)$ by solving a WLS multiple-regression problem with $2N$ observations, one response variable, vec(Y) and four carriers. The carriers correspond to regression parameters $\beta\cos\psi$, $\beta\sin\psi$, $\boldsymbol{\gamma}_1$ and $\boldsymbol{\gamma}_2$ (Goodall, 1983; Merickel, 1988). This approach is not useful for size-and-shape analysis. For general Σ and $K\geqslant 3$ the analogous regression is non-linear, with $K(K+1)/2+1$ parameters.

8. GENERALIZED PROCRUSTES ANALYSIS

GPA with $\Sigma=I$ is described by Gower (1975). We first outline his algorithm, with Ten Berge's (1977) modification, and then discuss the weighted case. GPA is an iterative algorithm, that successively updates the X_i' to minimize the generalized Procrustes sum of squares (2.4) or, equivalently,

$$\sum_{i<j}\|X_i'-X_j'\|^2 \tag{8.1}$$

subject to the constraint

$$\sum_i \mathrm{tr}(\tilde{X}_i'\tilde{X}_i'^{\mathrm{T}})=\sum_i \mathrm{tr}(\tilde{X}_i\tilde{X}_i^{\mathrm{T}}). \tag{8.2}$$

Estimates of the translations, rotations and scales of the similarity transformations $(\boldsymbol{\gamma}_i, \beta_i, \Gamma_i)$ (5.1) are computed in that order. The translation is handled once and for all by centring. The rotations Γ_i are determined iteratively. Starting with $X_i'^{(\mathrm{old})}=X_i$, let

$$\bar{X}_{(i)}=\frac{1}{L-1}\sum_{j\neq i}X_j'^{(\mathrm{old})}. \tag{8.3}$$

Then $X_i'^{(\mathrm{new})}$ is the ordinary Procrustes superimposition, involving only rotation, of $X_i'^{(\mathrm{old})}$ on $\bar{X}_{(i)}$. The L figures are rotated in turn. At each step expression (8.1) decreases, until at convergence the matrices $X_i'^{\mathrm{T}}\bar{X}_{(i)}$ are symmetric positive semidefinite. In practice 3–5 iterations, each containing L OPA steps, often suffice.

Lastly, we deal with scaling. Let Φ: $L\times L$ be the correlation matrix of the vec(X_i') with principal eigenvector ϕ corresponding to the largest eigenvalue. Then (Ten Berge, 1977)

$$\hat{\beta}_i=\{(\sum_k\|X_k'\|^2)/\|X_i'\|^2\}^{1/2}\phi_i. \tag{8.4}$$

But then the $\bar{X}_{(i)}$ may have changed and it is necessary to repeat the rotation iterations. The complete algorithm will include a total of between $3\times 2\times L$ and $5\times 3\times L$ OPA steps (Table 3), at each of which the sum of squares (8.1) does not increase.

OPA is, with some small adjustments (Gower, 1975) the generalized version with $L=2$. As already noted, the generalized Procrustes sum of squares is the sum of L Procrustes sums of squares. Some modifications, typically unimportant in practice,

TABLE 3

GPA rotations and scales for a sample of 10 undernourished rats, showing 8 rotation and scale steps

Rat	*Rotations (deg)*				*Scale*	*Rotations*		*Scale*
	Step 1	*Step 2*	*Step 3*	*Step 4*	*Step 5*	*Step 6*	*Step 7*	*Step 8*
1	0.04	−0.271	−0.0352	−0.00014	0.976	0.000015	−0.0000187	1
2	−1.33	−0.151	−0.0220	0.00233	0.987	0.000001	−0.0000208	1
3	0.85	−0.256	0.0026	0.00231	1.050	−0.000121	−0.0000097	1
4	−1.33	−0.150	0.0178	0.00075	1.100	−0.000099	0.0000002	1
5	−2.54	0.119	0.0064	0.00018	0.990	−0.000075	0.0000086	1
6	1.06	0.019	0.0051	−0.00042	1.030	0.000146	−0.0000067	1
7	−0.24	0.049	0.0001	−0.00048	0.973	0.000010	−0.0000085	1
8	0.73	−0.032	0.0039	−0.00100	0.941	−0.000094	0.0000010	1
9	−1.34	0.119	−0.0093	−0.00005	0.972	−0.000127	0.0000152	1
10	1.73	−0.059	−0.0037	0.00036	1.000	0.000191	−0.0000044	1

similar to those discussed for OPA, apply when GPA is used to estimate a mean shape, and not a mean reflection shape.

When $\Sigma \neq I$, $\Sigma = \Sigma_N \otimes \Sigma_K$, expression (8.1) becomes

$$\sum_{i<j} \| X_i' - X_j' \|_\Sigma^2 \tag{8.5}$$

and the generalized Procrustes sum of squares is

$$G_{\gamma,\beta,\Gamma}(\mathbf{X};\Sigma) = \sum_{i=1}^{L} G^*_{\gamma,\beta,\Gamma}(X_i, \bar{X};\Sigma) = \sum_{i=1}^{L} \| X_i' - \bar{X} \|_\Sigma^2. \tag{8.6}$$

where $\bar{X} = \Sigma X_i'/L$. The translation, rotation and scaling steps of GPA are modified as needed. From equations (7.10) the X_i can be projected separately, by $I-B$, in place of centring. The rotation steps are modified using either the Cholesky-transformed data, Koschat and Swayne's algorithm or its generalization when $K=2$. For the scaling step we replace each cross-product $\mathrm{vec}(X_i')^{\mathrm{T}}\mathrm{vec}(X_j')$ by $\mathrm{vec}(X_i')^{\mathrm{T}}\Sigma^{-1}\mathrm{vec}(X_j')$.

9. SUPERIMPOSITION METRIC Σ^{S} AND MODEL METRIC Σ^{M}

The population means in the one- and two-sample models (Section 6) are estimated using specific sequences of OPA and GPA steps. Before describing these procedures, we recognize that the metric Σ used in Procrustes analysis may differ from the metric Σ of the model. We call these the *superimposition metric* Σ^{S} and *model metric* Σ^{M} respectively. Any unknown parameters in the model metric must be estimated. The superimposition metric is selected by the user; it may equal the estimate of the model metric. $\hat{\Sigma}^{\mathrm{M}}$ makes sense, but $\hat{\Sigma}^{\mathrm{S}}$ does not.

A key intuition is that many choices of superimposition metric yield consistent estimates of the population means, or other parameters, such as regression parameters (Section 16). In model (5.1) the $(K(K+1)/2+1)$-parameter similarity transformations $(\omega_i, \alpha_i, \Omega_i)$ act on each landmark of the $\mu + E_i$ separately and can be estimated using any subset of at least $(K+1)/2+1/K$ landmarks in any metric (a submatrix of Σ^{S}). The resulting Procrustes estimate of the population mean may be consistent either as $L\to\infty$ or $\mathrm{tr}(\Sigma^{\mathrm{M}})\to 0$. Given an estimate $[\hat{\mu}]$ of $[\mu]$ we may estimate Σ^{M} from the cross-product of the residuals $X_i' - \hat{\mu}$ (Section 10). Confidence regions and

hypothesis testing should be based on the model metric and may be biased if the superimposition metric is used uncritically for inference. Four strategies for choosing the metrics are as follows.

(a) Use superimposition metric $\Sigma^S = I_{NK}$, allowing unweighted OPA and GPA. Goodall and Bose (1987) further assume that $\Sigma^M = I$; a better choice is to estimate Σ^M.
(b) For planar data, superimpose on an edge and base inference on the distribution of the resulting Bookstein shape co-ordinates (Bookstein, 1986; Bookstein and Sampson, 1990). This is a limiting case of our general framework, when $(\Sigma^S)^{-1}$ has infinite entries in the 4×4 submatrix corresponding to the two selected landmarks. The covariance of the remaining $2(N-2)$ shape co-ordinates is the estimate of Σ^M. This approach can be adapted to the other systems of shape co-ordinates mentioned in Sections 1 and 4.
(c) Use robust superimposition, which may involve different weighting of landmarks in different figures, and inference based on an estimate $\hat{\Sigma}^M$.
(d) Starting with an initial superimposition, $\Sigma^S = I$ say, the model metric is estimated iteratively from the data and used for superimposition, $\Sigma^S = \hat{\Sigma}^M$. Estimates and tests are efficient *when the two metrics coincide*, i.e. the iterations converge.

The details of maximum likelihood estimation, strategy (d), are given for the one- and two-sample models in the next section.

10. ITERATIVELY REWEIGHTED LEAST SQUARES FOR SHAPE DATA

10.1. *Efficient Estimation in One-sample Model*

A four-step strategy to compute MLEs for the one-sample model is the following:

(a) initial choice of Σ^S, $\Sigma^S = I$ say; set $\hat{\Sigma}^M = \Sigma^S$;
(b) GPA($\hat{\Sigma}^M$) to determine $\hat{\mu}$ and the X_i';
(c) estimation of the model covariance;
(d) OPA($\hat{\Sigma}^M$) to superimpose ν on $\hat{\mu}$.

Write $R_i = X_i' - \hat{\mu}$ and $\mathbf{r}_i = \mathrm{vec}(R_i)$. The MLE of the unrestricted model metric Σ^M in model (5.1) is

$$\hat{\Sigma}^M = \sum_i \mathbf{r}_i \mathbf{r}_i^T / L. \tag{10.1}$$

When Σ^M is factored model (5.3) the MLE decomposes as $\hat{\Sigma}_X^M = \hat{\Sigma}_N^M \otimes \hat{\Sigma}_K^M$, where

$$\hat{\Sigma}_K^M = \frac{1}{LN} \sum_i R_i^T R_i, \tag{10.2}$$

$$\hat{\Sigma}_N^M = \frac{1}{LK} \sum_i R_i R_i^T / \mathrm{tr}(\hat{\Sigma}_K). \tag{10.3}$$

The generalized Kronecker product covariance (5.4) is estimated similarly, with the proviso that we may take shrinkage estimates $\lambda \hat{\Sigma}_{Kj}^M + (1-\lambda)\hat{\Sigma}_K^M$ to avoid singularities due to rank $\hat{\Sigma}_{Kj}^M < 2$. When Σ^M is a multiple of the identity, $\Sigma^M = \epsilon^2 I$, the MLE is

$$\hat{\epsilon}^2 = \sum_i \mathbf{r}_i^{\mathrm{T}} \mathbf{r}_i / LKN. \tag{10.4}$$

In general, the structure of the model metric is chosen, and the appropriate MLE, is then computed from the R_i. Each $X_i' - \hat{\mu}$ is in $T_{\hat{\mu}}$, so unrestricted estimates $\hat{\Sigma}_M$ or $\hat{\Sigma}_N^{\mathrm{M}}$ will be singular, but nevertheless desirable (Section 5). These expressions for the MLE of Σ^{M} do not depend on Σ^{S}; however, the distributions of $\hat{\mu}$ and $\hat{\Sigma}^{\mathrm{M}}$ do depend on Σ^{S}. The joint likelihood either increases or does not change at both steps (b) and (c) (provided that the class of permitted covariance matrices remains the same or becomes larger). Steps (b) and (c) may be repeated to convergence.

10.2. *Efficient Estimation for Two-sample Model: Cross-sectional Data*

We assume that, when superimposition is complete, $\hat{\mu}_X' \approx \hat{\mu}_Y$ and also that the two model metrics are equal, $\Sigma_{X'}^{\mathrm{M}} = \Sigma_Y^{\mathrm{M}}$. To pool the covariance estimates we include the OPA superimposition, step (d), at each iteration:

(a) initial choice of Σ_X^{S} and Σ_Y^{S} to superimpose the X_i and Y_i respectively; set $\hat{\Sigma}_X^{\mathrm{M}} = \Sigma_X^{\mathrm{S}}$ and $\hat{\Sigma}^{\mathrm{M}} = \hat{\Sigma}_Y^{\mathrm{M}} = \Sigma_Y^{\mathrm{S}}$;
(b) GPA($\hat{\Sigma}_X^{\mathrm{M}}$) and GPA($\hat{\Sigma}_Y^{\mathrm{M}}$) to determine $\hat{\mu}_X$, X_i' and $\hat{\mu}_Y$, Y_i' respectively;
(c) estimation of the model covariances from the $X_i' - \hat{\mu}_X$ and $Y_i' - \hat{\mu}_Y$ respectively;
(d) OPA($\hat{\Sigma}^{\mathrm{M}}$) to superimpose $\hat{\mu}_X$ on $\hat{\mu}_Y$.

Compute $\Sigma_{X'}^{\mathrm{M}}$ using equation (6.4) and pool the covariances to give

$$\hat{\Sigma}^{\mathrm{M}} = (L_X \hat{\Sigma}_{X'}^{\mathrm{M}} + L_Y \hat{\Sigma}_Y^{\mathrm{M}})/(L_X + L_Y). \tag{10.5}$$

Set $\hat{\Sigma}_X^{\mathrm{M}} = \hat{\Sigma}_Y^{\mathrm{M}} = \hat{\Sigma}^{\mathrm{M}}$.

The joint likelihood of the two samples is non-decreasing at each step. Steps (b)–(d) are repeated to convergence. A modification of this algorithm to allow more than two samples, one-way ANOVA, is used in Section 14.

11. INFERENCE

Sibson (1979) shows that when $E \sim N(0, \epsilon^2 I)$ in model (7.2), where ϵ is small, then the ordinary Procrustes sums of squares $G^*_{\gamma,\beta,\Gamma}(X, Y)$ and $G^*_{\gamma,\Gamma}(X, Y)$ have central χ^2-distributions. These results are extended to GPA by Langron and Collins (1985). Perturbation analysis for GPA uses the first iteration only and is then essentially the same as perturbation analysis for OPA, with the multiplication of each degree of freedom by $L-1$. This follows because GPA can be regarded as multivariate ANOVA (MANOVA) in the Procrustes tangent space $T_{\hat{\mu}}$. Langron and Collins (1985) present ANOVA tables for both OPA and GPA, including sums of squares for rotation, translation and scale components of the similarity transformations. In the context of shape analysis, their ANOVA is appropriate for comparing alternative superimpositions but does not directly provide tests for shape differences.

As a foundation for multivariate analysis we derive a *multivariate Procrustes statistic*. The techniques used are similar to those of Sibson (1979). However, Sibson, and Langron and Collins, computed only aggregate expressions for the sums of squares.

Our derivation uses the SVD of X and is consequently more direct. Goodall and Bose (1987) sketch the derivation of a multivariate Procrustes statistic, without the SVD. With the roles of X and Y reversed here relative to these accounts, the (negated) residuals $R = X' - Y$ are orthogonal to X, not to Y. Both Procrustes tangent spaces $T_{X'}$ and T_Y contain R, up to scaling which is non-linear.

11.1. *Multivariate Procrustes Statistic,* $\Sigma^S = \Sigma^M = I_{NK}$

Let X and Y be a pair of figures, with X mean centred. Let X'_Γ, $X'_{\gamma,\Gamma}$, and $X' = X'_{\gamma,\beta,\Gamma}$ be the Procrustes fits of X to Y allowing respectively rotation (and reflection), rotation and translation, and rotation, translation and scale. We define the $N \times K$ multivariate Procrustes statistics S_Γ, $S_{\gamma,\Gamma}$ and $S = S_{\gamma,\beta,\Gamma}$ to be respectively the differences $X'_\Gamma - Y$, $X'_{\gamma,\Gamma} - Y$ and $R = X' - Y$ *double rotated*, in both $\mathbf{R}^N$ and $\mathbf{R}^K$, to the principal axes of the model $\beta X\Gamma$.

To include translation, we use the SVD of the mean-centred X instead of the SVD of the preshape Z (4.1). Assume that $N > K$, and write

$$X = \tilde{U}\Xi V^T, \tag{11.1}$$

where Ξ: $N \times K$ is zero apart from the diagonal elements $\xi_i > 0$, $V \in SO(K)$ and $\tilde{U}$: $N \times N \in O(N)$ with Nth column $\mathbf{1}_N/\sqrt{N}$, which is possible because X is mean centred. The vectors of singular values of Z (4.1) and X (11.1) are related by $\boldsymbol{\lambda} = \boldsymbol{\xi}/\sqrt{(\boldsymbol{\xi}^T\boldsymbol{\xi})}$, the Vs in equations (4.1) and (11.1) are identical and $U = H\tilde{U}$.

The multivariate Procrustes statistic for rotation only is

$$S_\Gamma = \tilde{U}^T(X'_\Gamma - Y)\Gamma^T V. \tag{11.2}$$

To first order in ϵ it is a simple linear function of the double-rotated errors,

$$P = \tilde{U}^T E\Gamma^T V. \tag{11.3}$$

Assume that $\beta = 1$ and $\gamma = 0$, i.e. X and the perturbed Y differ by a rotation Γ only. Then

$$\begin{aligned} S_\Gamma &= \tilde{U}^T(X\hat{\Gamma} - Y)\Gamma^T V \\ &= \tilde{U}^T\left[(X\{(X^T(X\Gamma - E)(X\Gamma - E)^T X\}^{-1/2}X^T - I_N\right](X\Gamma - E)\Gamma^T V \\ &= \left[\Xi\{\Xi^T(\Xi - P)(\Xi - P)^T\Xi\}^{-1/2}\Xi^T - I_N\right](\Xi - P). \end{aligned} \tag{11.4}$$

To find the inverse root, we equate the $K \times K$ term in braces to $(\Xi^T\Xi - \epsilon C)^2$, to first order in ϵ. This gives

$$\Xi^T\Xi P^T\Xi + \Xi^T P\Xi^T\Xi = \epsilon\Xi^T\Xi C + \epsilon C\Xi^T\Xi,$$

so that, writing $P = (P_{ij})$, $C = (C_{ij})$,

$$\epsilon C_{ij} = (\xi_j P_{ji}\xi_i^2 + \xi_i P_{ij}\xi_j^2)/(\xi_i^2 + \xi_j^2).$$

Equating

$$(\Xi^T\Xi - \epsilon C)\{(\Xi^T\Xi)^{-1} + \epsilon A\} = I,$$

then, to first order in ϵ,

$$A_{ij} = C_{ij}/\xi_i^2\xi_j^2.$$

A and C are both $K \times K$ matrices. Substitution into equation (11.4) gives, for rotation only,

$$(S_\Gamma)_{ij} = \begin{cases} P_{ij} + O(\epsilon^2), & i = j, \\ \dfrac{\xi_j^2 P_{ij} + \xi_i \xi_j P_{ji}}{\xi_i^2 + \xi_j^2} + O(\epsilon^2), & i > K, \\ P_{ij} + O(\epsilon^2), & 1 \leqslant i \neq j \leqslant K. \end{cases} \tag{11.5}$$

For rotation and translation,

$$(S_{\gamma,\Gamma})_{ij} = \begin{cases} (S_\Gamma)_{ij}, & i < N, \\ 0, & i = N. \end{cases} \tag{11.6}$$

For rotation, translation and scale,

$$S_{ij} = \begin{cases} (S_{\gamma,\Gamma})_{ij}, & i \neq j, \\ s_i, & i = j, \end{cases} \tag{11.7}$$

where, writing $\mathbf{p} = (p_i) = (P_{ii})$ and $\mathbf{s} = (s_i)$,

$$\mathbf{s} = \left(I - \frac{\boldsymbol{\xi}\boldsymbol{\xi}^{\mathrm{T}}}{\boldsymbol{\xi}^{\mathrm{T}}\boldsymbol{\xi}} \right) \mathbf{p}. \tag{11.8}$$

The distributional results of Sibson (1979) follow immediately from equations (11.5)–(11.8) and the distributional assumptions on E: asymptotically

$$\left.\begin{aligned} G_\Gamma^* = \| S_\Gamma \|^2 &\sim \epsilon^2 \chi^2_{NK - \frac{1}{2}K(K-1)} \\ G^*_{\gamma,\Gamma} = \| S_{\gamma,\Gamma} \|^2 &\sim \epsilon^2 \chi^2_{NK - \frac{1}{2}K(K+1)} \\ G^* = G^*_{\gamma,\beta,\Gamma} = \| S_{\gamma,\beta,\Gamma} \|^2 &\sim \epsilon^2 \chi^2_{NK - \frac{1}{2}K(K+1) - 1}. \end{aligned}\right\} \tag{11.9}$$

As a simple extension, we give the distribution of G^* when $P \sim N(P_0, \epsilon^2 I_{NK})$. The Procrustes fit removes the component similar to X from both P_0 and the random part $P - P_0$. Some algebra, or a little reflection, shows that to first order

$$G^*(X, Y) = \| S \|^2 \sim \epsilon^2 \chi^2_m \{ G^*(X, P_0) \}. \tag{11.10}$$

12. GEOMETRICAL INTERPRETATION AND RESIDUALS

In the double-rotated space, $\tilde{U}^{\mathrm{T}} X V = \Xi$ and $\tilde{U}^{\mathrm{T}} Y \Gamma^{\mathrm{T}} V = \Xi - P$. Any pair of columns of $\tilde{U}^{\mathrm{T}} X V$, the ith and jth say with $1 \leqslant i < j \leqslant K$, are orthogonal N-vectors, along the respective co-ordinate axes, with lengths ξ_i and ξ_j. In the (i, j) co-ordinate plane these differ from the corresponding Y-vectors by (P_{ii}, P_{ij}) and (P_{ji}, P_{jj}) respectively. To minimize the sum-of-squares error we rotate X in that plane through an angle ψ_{ij}. The minimum is achieved when

$$\tan \psi_{ij} = \frac{\xi_j P_{ji} - \xi_i P_{ij}}{\xi_i^2 + \xi_j^2 - \xi_i P_{ii} - \xi_j P_{jj}}. \tag{12.1}$$

To first order in ϵ the expression $\xi_i P_{ii} + \xi_j P_{jj}$ in the denominator of equation (12.1) can be ignored. We choose ψ_{ij} to minimize

$$(P_{ij} + \psi_{ij}\xi_i)^2 + (P_{ji} - \psi_{ij}\xi_j)^2, \tag{12.2}$$

and the choice of one ψ_{ij} does not affect the other ψ_{ij} (to first order). S_{ij} and S_{ji} then satisfy equation (11.5). The $K(K-1)/2$ pairwise ψ_{ij} together determine the 'rotation error' matrix $\hat{\Gamma}\Gamma^{\mathrm{T}}$. In double-rotated co-ordinates

$$(\hat{\Gamma}\Gamma^{\mathrm{T}})_{ij} = \delta_{ij} + \frac{\xi_j P_{ji} - \xi_i P_{ij}}{\xi_i^2 + \xi_j^2}, \tag{12.3}$$

a fact buried in equation (11.4). With β not necessarily unity, we choose the scale $\hat{\beta}$ to minimize

$$\sum_{i=1}^{K} (\xi_i - P_{ii} - \beta^{-1}\hat{\beta}\,\xi_i)^2, \tag{12.4}$$

which is regression through the origin. Thus

$$\beta^{-1}\hat{\beta} = 1 - \sum P_{ii}\xi_i / \sum \xi_i^2. \tag{12.5}$$

The residuals S_{ii} then satisfy equation (11.8), and the combined scale–rotation matrix $\beta^{-1}\hat{\beta}\Gamma^{\mathrm{T}}\hat{\Gamma}$ has diagonal elements (12.5) and off-diagonal elements (12.3).

From $S_{ji} = \xi_i/\xi_j S_{ij}$, $\Xi^{\mathrm{T}}S$ is symmetric, or equivalently $Y^{\mathrm{T}}X\hat{\Gamma}$ is symmetric. However, the *exact* residuals from fitting rotations do not satisfy these constraints with $\tan\psi_{ij}$ given by equation (12.1). This is because the double-rotated co-ordinates are updated by the rotations ψ_{ij}. When this adjustment is made, the exact double-rotated residuals S from the Procrustes superimposition (with rotation, translation and scale) of X on Y satisfy the $K(K+1)/2+1$ linear constraints found in the first-order expressions (11.5)–(11.8), namely

$$\left.\begin{aligned} \xi_i S_{ij} &= \xi_i S_{ji}, && 1 \leqslant i < j \leqslant K, \\ S_{Nj} &= 0, && 1 \leqslant j \leqslant K, \\ \sum_{i=1}^{K} \xi_i S_{ii} &= 0. \end{aligned}\right\} \tag{12.6}$$

These are the global, not just first-order, constraints that define the Procrustes tangent space (Section 3).

13. INFERENCE FOR GENERAL Σ^{S} AND Σ^{M}

Tests for difference in shape are generally based on a Hotelling's T^2-statistic in $T_{\hat{\mu}}$, written $\|R\|^2_{\hat{\Sigma}}$, where $\hat{\Sigma} = \hat{\Sigma}^{\mathrm{M}}$ is a covariance estimate from the R_i. Alternative forms of the statistic are $\mathrm{tr}(R^{\mathrm{T}}\hat{\Sigma}_N^{-1}R\hat{\Sigma}_K^{-1})$ when $\hat{\Sigma} = \hat{\Sigma}_N \otimes \hat{\Sigma}_K$ and $\|R\|^2/\hat{\epsilon}^2$ when $\hat{\Sigma} = \hat{\epsilon}^2 I_{NK}$.

For general metrics, we first examine the situation when the superimposition and model metrics coincide, $\Sigma^{\mathrm{S}} = \Sigma^{\mathrm{M}} = \Sigma$. When $\Sigma = \epsilon^2 I_{NK}$ then asymptotically $\|S\|^2$ has a χ^2-distribution by equations (11.9) and $\|S\|^2/\hat{\epsilon}^2$ has an F-distribution. Details for the one- and two-sample models follow in Section 14. For $\Sigma = \Sigma_N \otimes I$, $(\Sigma_N)^{-1} = Q^{\mathrm{T}}Q$, the same results, including equations (11.5)–(11.9), hold (premultiply

X and Y by Q and double rotate to the principal axes of QX). For more general Σ we may use the asymptotic properties of maximum likelihood and Hotelling's T^2-statistic: -2 log-likelihood, essentially $\|R\|^2_{\tilde{\Sigma}}$, asymptotically has a χ^2-distribution (cf. Anderson (1984), theorem 5.2.3). Alternative specifications of the model metric can also be compared using Wilks's likelihood ratio test.

Equations (11.5)–(11.8) show that the structure of the multivariate Procrustes statistic does not depend on the distribution of P. Thus when the metrics do not coincide we may be able to compute the distribution of the multivariate Procrustes statistic. (However, when $\Sigma^{\mathrm{S}}_K \neq I$ we do not have an explicit form for the rotation to insert into equation (11.4). This is unlikely to occur in practice because $\Sigma^{\mathrm{S}}_K = I$ is preferred for computation.) The most frequent metric mismatches are

$$\Sigma^{\mathrm{S}} = I_{NK}, \qquad \Sigma^{\mathrm{M}} = \Sigma_N \otimes \Sigma_K \neq I_{NK} \tag{13.1}$$

and

$$\Sigma^{\mathrm{S}} = \Sigma_N \otimes I_K, \qquad \Sigma^{\mathrm{M}} = \Sigma_N \otimes \Sigma_K. \tag{13.2}$$

Given equations (13.2) the distribution of the S is computed using equations (11.5)–(11.8) with $\operatorname{cov}\{\operatorname{vec}(P)\} = \epsilon^2(\tilde{U}^{\mathrm{T}}Q \otimes V^{\mathrm{T}})\Sigma^{\mathrm{M}}(Q^{\mathrm{T}}\tilde{U} \otimes V)$.

As a further example, we express the Procrustes statistic $G^*(X, Y)$ as a quadratic form when $K = 2$, $\Sigma^{\mathrm{S}} = I_{NK}$, and Σ^{M} has the generalized factored form (5.4). Then

$$\operatorname{vec}(S) = ABC\mathbf{u} \tag{13.3}$$

where $\mathbf{u} \sim N(0, I_{NK})$ and A, B and C are each $NK \times NK$ matrices of constants, $CC^{\mathrm{T}} = \Sigma^{\mathrm{M}}$, $B^{\mathrm{T}} = \tilde{U} \otimes V$ and A is block diagonal. The three blocks of A are, from equations (11.5)–(11.8), A^*: 4×4, $I_{2(N-3)}$, and a 2×2 matrix of zeros, where A^* appears in

$$\begin{pmatrix} S_{11} \\ S_{12} \\ S_{21} \\ S_{22} \end{pmatrix} = \frac{1}{\xi_1^2 + \xi_2^2} \begin{pmatrix} \xi_2^2 & 0 & 0 & -\xi_1\xi_2 \\ 0 & \xi_2^2 & \xi_1\xi_2 & 0 \\ 0 & \xi_1\xi_2 & \xi_1^2 & 0 \\ -\xi_1\xi_2 & 0 & 0 & \xi_1^2 \end{pmatrix} \begin{pmatrix} P_{11} \\ P_{12} \\ P_{21} \\ P_{22} \end{pmatrix}. \tag{13.4}$$

A^* is a highly structured, symmetric idempotent matrix that is familiar in Procrustes analysis with $K = 2$. It may be called the *Procrustes projection* matrix.

Let $C_{Kj}C_{Kj}^{\mathrm{T}} = \Sigma_{Kj}$ be a decomposition of the individual landmark covariance and Q be as in equation (5.4). The Procrustes statistic is

$$G^* = \|S\|^2 = \mathbf{u}^{\mathrm{T}} \oplus C_{Kj}^{\mathrm{T}}(Q^{-\mathrm{T}}\tilde{U} \otimes V)A(\tilde{U}^{\mathrm{T}}Q^{-1} \otimes V^{\mathrm{T}}) \oplus C_{Kj}\mathbf{u}. \tag{13.5}$$

When $\Sigma^{\mathrm{M}} = I_{NK}$ the standard central χ^2-result (11.9) follows immediately. Otherwise we may approximate G^* by a scaled χ^2-distribution by matching first and second moments. For example,

$$EG^* = \operatorname{tr}\{A(\tilde{U}^{\mathrm{T}}Q^{-1} \otimes V^{\mathrm{T}}) \oplus \Sigma_{Kj}(Q^{-\mathrm{T}}\tilde{U} \otimes V)\}. \tag{13.6}$$

In related work Davis (1978) derives a scaled χ^2-approximation to the distribution of the Procrustes statistic (Gower's m^2) in superimposing two configurations X and

Y with $\Sigma^{\mathrm{M}} = I_N \otimes \Sigma_K$, where before OPA each configuration is *transformed to its canonical variates*, e.g.

$$X^* = X\hat{\Sigma}_K^{-1/2}, \tag{13.7}$$

where $\hat{\Sigma}_K$ is the estimated covariance (10.2). Then X^* and Y^* have covariance I_{NK}, to first order. This is a different set-up from shape analysis, where superimposition must use similarity transformations in the original space, not the canonical variable space. $\hat{\Sigma}_K^{-1/2}$ does not commute with $V \in O(K)$, even when $K=2$. Sibson (1979) includes some further perturbation analysis of Davis's (and related) problems.

14. TESTS FOR SHAPE DIFFERENCES

Tests for the one-sample and two-sample models are based on the inference results of Section 11. We discuss the extension to one-way ANOVA and apply the tests to some nutritional data.

14.1. *One-sample Test for Shape Data*

In the one-sample case, with $\Sigma^{\mathrm{S}} = \Sigma^{\mathrm{M}} = \epsilon^2 I_{NK}$, the two Procrustes sums of squares have independent distributions $G^*(\nu, \hat{\mu}) \sim \epsilon^2 / L\chi^2_m$ and $G(\mathbf{X}) \sim \epsilon^2 \chi^2_{(L-1)m}$. The F-statistic for testing H_0: $[\mu] = [\nu]$ against general alternatives is

$$(L-1)L\frac{G^*(\nu, \hat{\mu})}{G(\mathbf{X})} \sim F_{m,(L-1)m} \tag{14.1}$$

14.2. *Two-sample Test for Shape Data*

In the two-sample case, with $\hat{\mu}_{X'} \approx \hat{\mu}_Y$ and $\Sigma^{\mathrm{S}} = \Sigma^{\mathrm{M}}_{X'} = \Sigma^{\mathrm{M}}_Y = \epsilon^2 I_{NK}$ (Section 10), the three Procrustes sums of squares have independent distributions, namely $G^*(\hat{\mu}_X, \hat{\mu}_Y) \sim \epsilon^2 / L\chi^2_m$, $G(\mathbf{X}') \sim \epsilon^2 \chi^2_{(L_X-1)m}$ and $G(\mathbf{Y}) \sim \epsilon^2 \chi^2_{(L_Y-1)m}$. Here $\mathbf{X}' = \beta\mathbf{X}$ (Section 6). The F-statistic for testing H_0: $[\mu_X] = [\mu_Y]$ against general alternatives is

$$\frac{L_X + L_Y - 2}{L_X^{-1} + L_Y^{-1}} \frac{G^*(\hat{\mu}_X, \hat{\mu}_Y)}{G(\mathbf{X}') + G(\mathbf{Y})} \sim F_{m,(L_X+L_Y-2)m}. \tag{14.2}$$

14.3. *One-way Analysis of Variance*

One-way ANOVA is a simple extension of the two-sample model from two to $M \geqslant 2$ groups of figures, $\mathbf{X}_j = (X_{1j}, \ldots, X_{1L})$ for $j = 1, \ldots, M$. Write $\hat{\boldsymbol{\mu}} = (\hat{\mu}_1, \ldots, \hat{\mu}_M)$. Step (d) in Section 10 uses GPA, and statistic (14.2) becomes

$$L(L-1)MG(\hat{\boldsymbol{\mu}})/(M-1)\sum_{j=1}^{M} G(\mathbf{X}_j') \sim F_{(M-1)m, M(L-1)m}, \tag{14.3}$$

where $\mathbf{X}_j'$ is $\mathbf{X}_j$ scaled by equation (8.4). The discrepancy of a factor of 2 between statistics (14.2) and (14.3) with $M=2$ occurs because $G^*(X, Y) = 2G(X, Y)$. The groups have equal sizes for simplicity; otherwise the GPA of means involves unequal weights. In pooling the covariances we assume that the $[\mu_j]$ are close together.

Two-way ANOVA requires that this last condition holds for both extents of the matrix of shapes. When this is met, two-way ANOVA becomes MANOVA in $T_{\bar{\mu}}$, where $\bar{\mu}$ is the grand mean.

14.4. *Shape Differences Associated with Nutrition*

We analyse landmark data from radiographs of the lateral (sagittal) sections of close-bred, pre-weaned male rats 21 days old (Moss *et al.*, 1987), with $N=20$ landmarks (Fig. 1) digitized for each of $L=10$ rats in each of $M=4$ dietary regimens. These regimens, imposed on the mother, are control, undernourished, undernourished with protein supplement and undernourished with carbohydrate supplement. The substantive question addressed is 'are differences in shape associated with differences in dietary regimen?'. The apparent biological mechanism is that differences in maternal diet result in differential rates of growth in the regions of the young rat head, which in turn result in different skull geometries. We compare the four samples using the two-sample and one-way ANOVA models.

Table 3 gives details of the generalized Procrustes fit, GPA(I_{NK}), for the undernourished sample. Fig. 5 plots the X_i' and shows that the covariance differs between landmarks. We fit the undernourished sample mean using the three covariances I_{NK}, $I_N \otimes \Sigma_K$, and $\oplus \Sigma_{Kj}$. Comparing the three metrics, there are some differences in the superimposition of individual rats, and the sums of squares differ by 2–4%; for example $G(X; I_{NK})$ is 3.5% larger than $G(X; \oplus \Sigma_{Kj})$. However, the differences between the three means are negligible: the Procrustes sum of squares comparing two means is typically 10^{-5} compared with GPA sums of squares around 200.

After computing each sample mean the next step is visualization of differences between regimens. With these data the OPA superimpositions turn out to be effective. The undernourished rats without supplement tend to have a taller neural skull than the control rats. The difference between the protein and carbohydrate supplements are localized at either extreme of the skull. The F-ratios comparing treatment to controls are 16.2, 17.3 and 7.4; compared with an F-distribution on 36 and 648 degrees of freedom each is highly significant. The one-way ANOVA F-ratio for all four regimens is 4.9, on 108 and 1296 degrees of freedom, also highly significant. We turn now to description of shape differences.

15. DESCRIPTION AND GEOMETRICAL COMPONENTS OF SHAPE DIFFERENCE

The description of a difference in shape as an affine transformation from one figure to the other is widespread: Bookstein (1984, 1986); Goodall (1983, 1984); Goodall and Green (1986); Mardia and Dryden (1989b); Rohlf and Slice (1990). Bookstein and Sampson (1990) give estimates and tests for linear (shear) and quadratic components of shape change. Goodall and Mardia (1990a) show that with the Gaussian model the marginal density of equivalence classes of figures modulo affine transformations is multivariate non-central beta.

Let A: $K \times K$ and $X' = X'_{\gamma, A}$ be the multivariate least squares regression fit of X to Y. A multivariate affine statistic $S_{\gamma, A}$ is defined to be $X'_{\gamma, A} - Y$ double rotated to the principal axes of $\beta X \Gamma$. Then

$$(S_{\gamma, A})_{ij} = \begin{cases} 0, & i \leqslant K, \\ S_{ij}, & i > K, \end{cases} \tag{15.1}$$

where S is the multivariate Procrustes statistic (11.7) for similarity transformations.

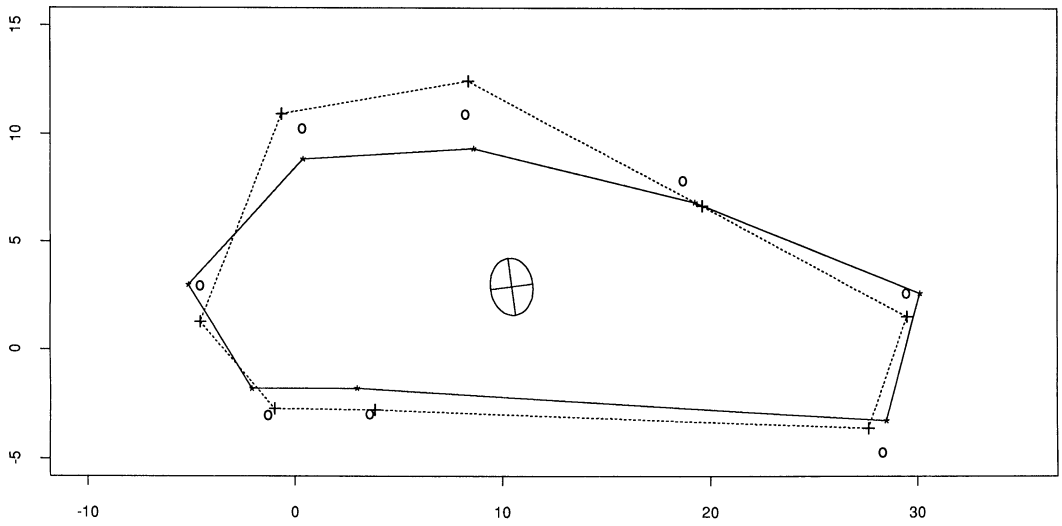

Fig. 6. Procrustes (∗) and affine (○) superimposition of the control mean on the hydrocephaly mean (+): the affine transformation includes perpendicular extensions by factors 1.25 and 0.96, shown by the strain cross and ellipse

$S_{\gamma, A}$ is orthogonal to X and is the usual matrix of multivariate regression residuals. This result establishes a hierarchy of models and tests for systematic differences in shape that bridges the ANOVA of Langron and Collins (1985) and conventional MANOVA model selection. Shape comparison is the zeroth-order component in a hierarchy of deformations: Euclidean/similarity (equivalent to form/shape), affine, quadratic, ... Apart from the non-linear zeroth-order component these are exact multivariate linear model results for Gaussian errors.

An affine transformation (Goodall, 1983, 1984, 1986; Goodall and Green, 1986) is interpreted, via a pair of strain crosses in the figures before and after deformation, as a rotation, followed by different strains in the co-ordinate directions, and a final rotation. To enhance interpretability the figures are superimposed, so that the $K \times K$ affine matrix is approximately symmetric and the original and final rotations are complementary. For the hydrocephaly data, the control mean is first superimposed on the hydrocephaly mean by OPA and then an additional affine transformation is computed (Fig. 6). The affine fit involves an extension of 25% at angle $-82.4°$ (before) and $-84.4°$ (after) and a contraction of 4% in the perpendicular direction. The decrease in the sum of squares, 12.8, is compared with an affine sum of squares, 11.9, by an F-ratio on two and 10 degrees of freedom. The F-ratio, 5.36, has the moderate P-value of 2.6%. The affine fit explains part of the difference between the control and hydrocephaly means, but Fig. 4 contains considerable additional graphical information.

16. DISCUSSION

Procrustes methods for the statistical analysis of shape are seen to be a viable approach to a highly structured but non-linear problem of estimation and inference. Inevitably in a new field, the paper leaves many questions only partly answered; we consider some of these.

16.1. *Large Shape Differences and Regression Models*

Cogent discussion of the relationships among shapes that are well separated requires some global structure on shape spaces. There are (at least) three approaches to consider.

(a) The affine transformations of Section 5 are sometimes useful, but more complex, or more local, transformations are also needed.
(b) Specific morphometric features may be of sufficient interest that we linearize that projection of shape space; for example in a study of the relative orientation of neural skull and facial skull in rats the *angle* between two major structures has primary interest.
(c) A regression model allows shapes to be compared along the regression curve.

A regression model for size-and-shape is a generalization of model (5.1). For figures X_i, $i=1, \ldots, L$, and a scalar carrier taking values x_i, the model is

$$X_i=(\mu+x_i\delta+E_i)\Omega_i+\mathbf{1}_N\omega_i^{\mathrm{T}} \tag{16.1}$$

where the E_i are random displacements, and μ and δ are respectively the 'intercept' and 'slope' matrices for $N\times K$ simple linear regressions coupled by the Euclidean transformations (ω_i, Ω_i). This model is most attractive when the size-and-shapes of $\mu+x\delta$, $x\in\mathbf{R}$, follow a geodesic in size-and-shape space, so $\mu+x\delta$ is a horizontal geodesic in preform space. Goodall (1990b) shows that the condition for this is that $\mu^{\mathrm{T}}\delta$ is symmetric but suggests that the unrestricted model is also important.

A regression model for shape may be written in terms of preshapes Z_i and an angular variable θ_i:

$$Z_i=(\mu\cos\theta_i+\mu^*\sin\theta_i+E_i)\Omega_i \tag{16.2}$$

where $\|\mu\|=\|\mu^*\|=1$, $\mathrm{tr}(\mu^{\mathrm{T}}\mu^*)=0$ and E_i is a Fisher disturbance (so $\|Z_i\|=1$). The geodesic in preshape space is horizontal when $\mu^{\mathrm{T}}\mu^*$ is symmetric, and again the unrestricted model is important. The regression model for shape without standardization for scale is obtained by including a term α_i in equation (16.1).

Goodall (1990b) fits regression models for a variety of data, including cases where the x_i or θ_i are unknown, or are unknown low order polynomial functions of a metameter t_i. A related method is Procrustes analysis of longitudinal data, sampled at two or more occasions (Goodall and Bose, 1987).

16.2. *Shape versus Size-and-shape*

Size by itself is a gross, if convenient, summary of a figure. Differences in shape are more subtle and multidimensional. However, a figure changes shape when it experiences growth that is either non-isotropic or inhomogeneous. Thus both size and shape should be included in a physical model. For example, regression model (16.1) for size-and-shape appears to be a better physical model than regression model (16.2) for shape alone. However, the statistical arguments of Mosimann (1970), Sampson and Siegel (1985) and others suggest that shape analysis is valuable even when differences in size *and* shape are important. Procrustes theory for size-and-shape is closely parallel to that developed here, but somewhat simpler as the preform geometry is Euclidean, instead of the spherical geometry of preshape space.

16.3. *Robust Estimation*

A Procrustes approach to robust estimation involves a choice of weight matrix $W=\hat{\Sigma}_N^{-1}$, where

$$W = \text{diag}\{w_1^2, \ldots, w_N^2\} \tag{16.3}$$

and w_i^2 is a robustness weight for the ith landmark based on the lengths of the residual vectors. Other robustness issues include bounded influence (of distant landmarks) and extensions to GPA, in which an entire figure and not simply a single landmark may be an outlier. These methods complement the techniques of Siegel and Benson (1982) and Rohlf and Slice (1990) based on the repeated median.

16.4. *Comparison of Approaches*

The simplest approach to shape analysis is to assume that shapes have a spherical Gaussian distribution in Bookstein co-ordinates. This is a first-order approximation to the exact shape distribution obtained from the Gaussian model for landmark data with isotropic errors. Using the exact distribution the mean shape can be estimated by maximum likelihood, but inference is handled by first-order methods, such as Wilks's λ. Procrustes methods involve starting from the same Gaussian model for landmark data, but treating the similarity transformations as nuisance parameters. A different set of MLEs results and inference is again first order.

There is at present no detailed data to help to choose between the marginal and full likelihood approaches. In practical terms, the exact shape densities are relatively heavy going, whereas Procrustes methods are more algorithmic. Procrustes methods extend relatively easily to more complicated models, such as correlated landmarks, non-isotropic errors, ANOVA, longitudinal data and regression. These extensions recognize the non-linearity of shape spaces: the MLE of the mean shape $[\mu]$ is computed exactly even from widely dispersed data (although the Gaussian model for landmark data may become questionable), but in a two-sample comparison (or more) the mean shapes should be sufficiently close to pool the covariances. In ANOVA *all* the subsample mean shapes must be close together. Similarly, the shapes must be concentrated for the analysis of paired data.

When fewer than N landmarks are used for superimposition, the estimates of $(\gamma_i, \beta_i, \Gamma_i)$ are less precise, and the sum of squares about the mean may increase markedly (Fig. 3). Intuitively, with N landmarks there are $N/2$ independent estimates of the $(\gamma_i, \beta_i, \Gamma_i)$ using edge superimposition. However, the differences between superimposition methods are second order only: to first order, superimposition removes the geometrical structure of each figure and allows it to be analysed as an m-vector in the Procrustes tangent space. With the data thus arranged as an $L \times m$ matrix, the mean is simply the column averages. However, the tangent spaces differ only to second order, and, by orthogonality, the column averages are the same whether row effects have been removed using all or only some landmarks.

More complex issues arise when the covariance is estimated from the data. Procrustes methods provide MLEs and first-order inference as usual, but their relative merit compared with simply estimating the covariance of Bookstein's shape co-ordinates from data is not known.

16.5. *Concluding Remarks*

Shape spaces Σ_K^N for figures other than triangles are complicated. We have shown that practical statistical models and techniques for the statistical analysis of shape data can be constructed in Euclidean space. The approach of this paper is to work

in figure space, with nuisance parameters for unknown similarity transformations. In place of a metric directly on shape space we use the Procrustes sum of squares between figures. For both the hydrocephaly and the nutrition data, the choice of superimposition metric does not appear crucial, but for efficient estimation and testing we prefer GPA in which the model metric coincides with the superimposition metric.

ACKNOWLEDGEMENTS

The author wishes to thank Robin Sibson for directing him to Procrustes methods for shape analysis, John Kent for detailed and insightful comments, Melvin Moss for providing the data, and Anjana Bose, Ian Dryden, David Kendall, Subhash Lele, Kanti Mardia and Geoffrey Watson for important discussions. This paper was prepared in part in connection with research at Princeton University sponsored by the Army Research Office (Durham), grant DAAL03-88-K-0045, and was in part aided by award 1 P01 HD 19446 from the National Institute of Child Health and Development to Columbia University, and additional support from the Science and Engineering Research Council and Leeds University.

REFERENCES

Anderson, T. W. (1984) *An Introduction to Multivariate Statistical Analysis*, 2nd edn. New York: Wiley.

Bookstein, F. L. (1978) The measurement of biological shape and shape change. *Lect. Notes Biomath.*, **24**.

——— (1984) A statistical method for biological shape comparisons. *J. Theor. Biol.*, **107**, 475–520.

——— (1986) Size and shape spaces for landmark data in two dimensions (with discussion). *Statist. Sci.*, **1**, 181–242.

——— (1989) Comment on A survey of the statistical theory of shape (by D. G. Kendall). *Statist. Sci.*, **4**, 99–105.

——— (1991) *Morphometric Tools for Landmark Data: Geometry and Biology*. Cambridge: Cambridge University Press. To be published.

Bookstein, F. L. and Reyment, R. A. (1989) Microevolution in Miocene *Brizalina* (Foraminifera) studied by canonical variate analysis and analysis of landmarks. *Bull. Math. Biol.*, **51**, 657–679.

Bookstein, F. L. and Sampson, P. D. (1990) Statistical models for geometric components of shape change. *Communs Statist. Theory Meth.*, **19**, 1939–1972.

Carne, T. K. (1990) The geometry of shape spaces. *Proc. Lond. Math. Soc.*, **61**, 407–432.

Chow, Y., Grenander, U. and Keenan, D. M. (1988) HANDS, a pattern theoretic study of biological shapes. *Technical Report*. Division of Applied Mathematics, Brown University, Providence.

Cox, D. R. and Snell, E. J. (1968) A general definition of residuals (with discussion). *J. R. Statist. Soc.* B, **30**, 248–275.

Crampin, M. and Pirani, F. A. E. (1986) Applicable differential geometry. *Lond. Math. Soc. Lect. Note Ser.*, **59**.

Davis, A. W. (1978) On the asymptotic distribution of Gower's m^2 goodness-of-fit criterion in a particular case. *Ann. Inst. Statist. Math.*, **30**, 71–79.

Dryden, I. L. (1989) The statistical analysis of shape data. *PhD Thesis*. Department of Statistics, University of Leeds.

Dryden, I. L. and Mardia, K. V. (1991) General shape distributions in a plane. *Adv. Appl. Probab.*, to be published.

Golub, G. H. and Van Loan, C. F. (1983) *Matrix Computations*. Baltimore: Johns Hopkins University Press.

Goodall, C. R. (1983) The statistical analysis of growth in two-dimensions. *PhD Dissertation*. Department of Statistics, Harvard University.

——— (1984) The growth of a two-dimensional figure: strain crosses and confidence regions. *Proc. Statist. Comput. Sect. Am. Statist. Ass.*, 165–169.

——— (1986) Comment on Size and shape spaces for landmark data (by F. L. Bookstein). *Statist. Sci.*, **1**, 181–242.
——— (1990a) Eigenshape analysis of a cut-grow mapping for triangles, and its application to phyllotaxis in plants. *SIAM J. Appl. Math.*, to be published.
——— (1990b) Regression models for shape and form. To be published.
Goodall, C. R. and Bose, A. (1987) Procrustes techniques for the analysis of shape and shape change. In *Computer Science and Statistics* (ed. R. M. Heiberger), pp. 86–92. Alexandria: American Statistical Association.
Goodall, C. R. and Green, P. B. (1986) Quantitative analysis of surface growth. *Bot. Gaz.*, **147**, 1–15.
Goodall, C. R., Lange, N. and Moss, M. L. (1990) Growth-curve models for repeated triangular shapes. To be published.
Goodall, C. R. and Mardia, K. V. (1990a) Multivariate aspects of shape theory. Submitted to *Ann. Statist.*
——— (1990b) A geometrical derivation of the shape density. *Adv. Appl. Probab.*, to be published.
Gower, J. C. (1970) Statistical methods of comparing different multivariate analyses of the same data. In *Mathematics in the Archaeological and Historical Sciences*. Edinburgh: Edinburgh University Press.
——— (1975) Generalized Procrustes analysis. *Psychometrika*, **40**, 33–50.
——— (1984) Multivariate analysis: ordination, multivariate scaling and allied topics. In *Handbook of Applicable Mathematics* (ed. W. Ledermann). New York: Wiley.
Huxley, J. (1932) *Problems of Relative Growth*, 2nd edn. New York: Dover Publications.
Jupp, P. E. (1988) Residuals for directional data. *J. Appl. Statist.*, **15**, 137–147.
Kendall, D. G. (1984) Shape-manifolds, procrustean metrics and complex projective spaces. *Bull. Lond. Math. Soc.*, **16**, 81–121.
——— (1985) Exact distributions for shapes of random triangles in convex sets. *Adv. Appl. Probab.*, **17**, 308–329.
——— (1989) A survey of the statistical theory of shape (with discussion). *Statist. Sci.*, **4**, 87–120.
Kendall, D. G. and Kendall, W. S. (1980) Alignments in two-dimensional random sets of points. *Adv. Appl. Probab.*, **12**, 380–424.
Kendall, D. G. and Le, H.-L. (1990) The Riemannian structure of Euclidean shape spaces. Submitted to *Ann. Statist.*
Kendall, W. S. (1990) The diffusion of Euclidean shape. In *Disorder in Physical Systems* (eds D. Welsh and G. Grimmett), pp. 203–217. Oxford: Oxford University Press.
Kent, J. T. (1990) The complex Bingham distribution and shape analysis. *S. S. Wilks Workshop on Shape Theory, Princeton.*
Koschat, M. A. and Swayne, D. (1989) A weighted Procrustes criterion. *Psychometrika*, to be published.
Langron, S. P. and Collins, A. J. (1985) Perturbation theory for generalized Procrustes analysis. *J. R. Statist. Soc.* B, **47**, 277–284.
Le, H.-L. (1990a) On geodesics in Euclidean shape spaces. *J. Lond. Math. Soc.*, to be published.
——— (1990b) A stochastic calculus approach to the shape distribution of k labelled independent $N(\mu_j, \sigma^2 I)$-random points. *Proc. Camb. Phil. Soc.*, to be published.
Lord, E. A. and Wilson, C. B. (1984) *The Mathematical Description of Shape and Form*. New York: Wiley.
Mardia, K. V. (1989a) Shape analysis of triangles through directional techniques. *J. R. Statist. Soc.* B, **51**, 449–458.
———(1989b) Comment on A survey of the statistical theory of shape (by D. G. Kendall). *Statist. Sci.*, **4**, 108–111.
Mardia, K. V. and Dryden, I. L. (1989a) The statistical analysis of shape data. *Biometrika*, **76**, 271–282.
——— (1989b) Shape distributions for landmark data. *Adv. Appl. Probab.*, **21**, 742–755.
Mardia, K. V., Kent, J. T. and Bibby, J. M. (1979) *Multivariate Analysis*. London: Academic Press.
Merickel, M. (1988) 3D reconstruction: the registration problem. *Comput. Vis. Graphics Img. Process.*, **42**, 206–219.
Mosier, C. I. (1939) Determining a simple structure when loadings for certain tests are known. *Psychometrika*, **4**, 149–162.
Mosimann, J. E. (1970) Size allometry: size and shape variables with characterizations of the lognormal and generalized gamma distributions. *J. Am. Statist. Ass.*, **65**, 930–948.
Moss, M. L., Pucciarelli, H. M., Moss-Salentijn, L., Skalak, R., Bose, A., Goodall, C. R. and Sen, K. (1987) Effects of pre-weaning undernutrition on 21 day-old male rat skull form as described by the finite element method. *Gegenb. Morph. Jb*, **133**, 837–868.

Regalia, P. A. and Mitra, S. K. (1989) Kronecker products, unitary matrices and signal processing applications. *SIAM Rev.*, **31**, 586–613.
Robert, P. and Escoufier, Y. (1976) A unifying tool for linear multivariate statistical methods: the *RV*-coefficient. *Appl. Statist.*, **25**, 257–265.
Rohlf, F. J. and Slice, D. (1990) Methods for comparison of sets of landmarks. *Syst. Zool.*, **39**, 40–59.
Sampson, P. D. and Siegel, A. F. (1985) The measure of 'size' independent of 'shape' for multivariate lognormal populations. *J. Am. Statist. Ass.*, **80**, 910–914.
Seber, G. A. F. (1984) *Multivariate Observations*. New York: Wiley.
Sibson, R. (1978) Studies in the robustness of multidimensional scaling: Procrustes statistics. *J. R. Statist. Soc.* B, **40**, 234–238.
——— (1979) Studies in the robustness of multidimensional scaling: perturbational analysis of classical scaling. *J. R. Statist. Soc.* B, **41**, 217–229.
Siegel, A. F. and Benson, R. H. (1982) A robust comparison of biological shapes. *Biometrics*, **38**, 341–350.
Small, C. G. (1981) Distributions of shape and maximal invariant statistics. *PhD Thesis*. University of Cambridge.
——— (1984) A classification theorem for planar distributions based on the shape statistics of independent tetrads. *Math. Proc. Camb. Phil. Soc.*, **96**, 543–547.
——— (1988) Techniques of shape analysis on sets of points. *Int. Statist. Rev.*, **56**, 243–257.
——— (1989) Comment on A survey of the statistical theory of shape (by D. G. Kendall). *Statist. Sci.*, **4**, 105–108.
Ten Berge, J. M. F. (1977) Orthogonal procrustes rotation for two or more matrices. *Psychometrika*, **42**, 267–276.
Watson, G. S. (1986) The shape of a random sequence of triangles. *Adv. Appl. Probab.*, **18**, 156–169.

DISCUSSION OF THE PAPER BY GOODALL

D. G. Kendall (University of Cambridge): This is the first time that the Society has turned its attention to shape theory, and we are lucky to have such a solid contribution from Colin Goodall. As it is rather complicated I shall not go through the paper in detail, and happily it fits in better with my inclinations to describe a recent development illustrating a fruitful rule of procedure: *let the geometry do all the work so far as is reasonably possible.*

I begin with the problem of deciding whether a given set of points in the plane contains 'too many near collinearities'. Fellows will recall a famous paper on this topic by Broadbent (1980). There for the most part it was three-point collinearities that were studied, with an angular criterion for near collinearity. That angular criterion can behave very badly when two or more points are close together, and moreover it has never been clear how it ought to be generalized when the points are taken k at a time and k exceeds 3.

I here present a new method that is free from both disadvantages and outline its use when assessing four-point collinearities in the plane. The extension to k points is straightforward, though the graphics are then more complicated. The extension to three or more dimensions presents new features that I shall not enlarge on here.

The shape of four not wholly coincident labelled points in the plane is represented by a point in a certain complex projective shape space that is four dimensional, and when the four points are collinear then the shape lies in a two-dimensional submanifold that is the projective 2-sphere obtained by identifying antipodal points on an ordinary sphere of radius 1.

Recently my colleague Hui-ling Le has determined the global geodesic geometry in the shape space Σ_m^k for k points in m dimensions, making a special study of the cut-locus phenomena, and also of the *geodesic* projection from an arbitrary shape onto the submanifold $\mathscr{D}$ of dimensionally deficient shapes ('collinear' shapes when the basic dimension m is 2). Thus, in the example considered here, $\mathscr{D}$ is the projective 2-sphere.

On specializing her general theorems to the present problem we are led to a natural numerical measure of non-collinearity that I shall call L. This is a simple increasing function of the length of the *geodesic* perpendicular dropped from the current shape point in the shape space onto the submanifold $\mathscr{D}$, and

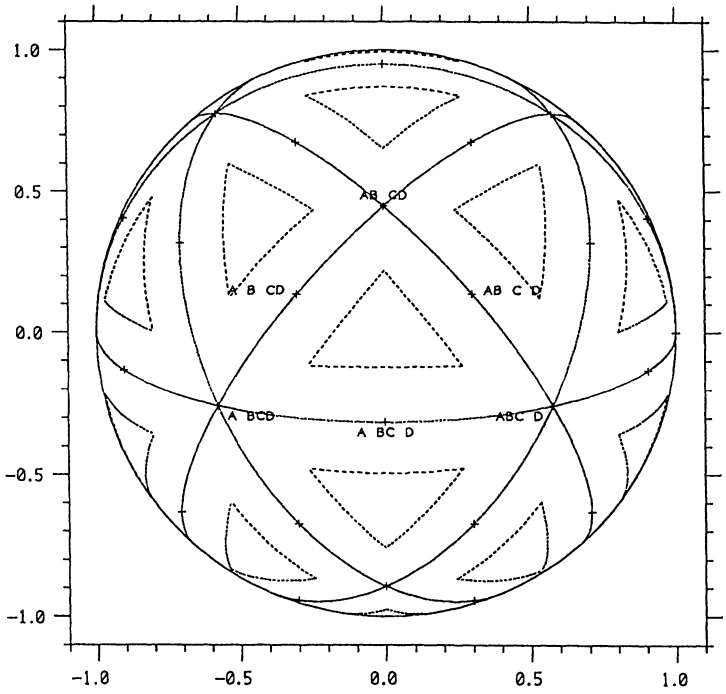

Fig. 7. RP^2 with tiles

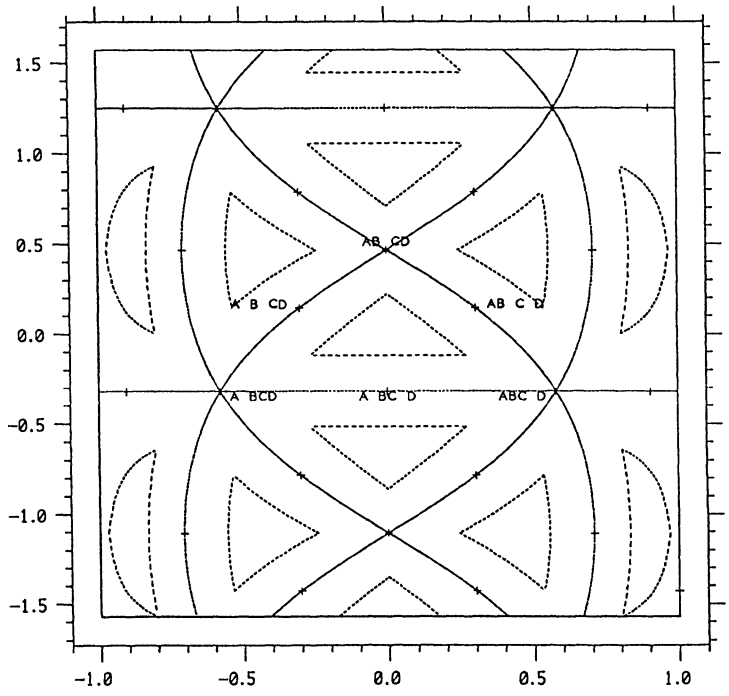

Fig. 8. RP^2 with tiles projected onto the cylindrical sleeve

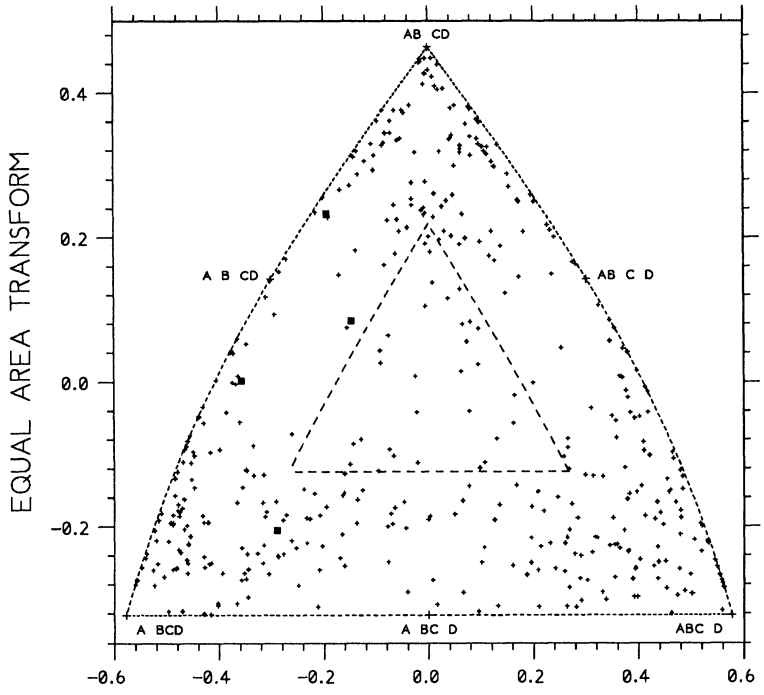

Fig. 9. Collinearity evidence displayed: ■, 0.0–0.00001; +, 0.00001–0.00100

associated with it there is an ancillary statistic M that locates the position of the foot of that geodesic perpendicular in the manifold $\mathscr{D}$, thus determining the nearest approximating shape that is dimensionally deficient. So in the present context L tells us how far from collinearity the tetrad is, while M gives the best collinear approximation.

My colleague has proved that, when the k points are independently and identically distributed Gaussian in the plane, then

(a) L has the uniform distribution on the unit segment,
(b) M has the differential geometric uniform distribution on the submanifold $\mathscr{D}$ of approximating collinear k-ads and
(c) these two statistics are mutually independent.

Thus we can regard the observed L as the appropriate test statistic for the collinearity problem, while M is an important and geometrically natural ancillary. Let us see what happens when we apply these ideas to the analysis of tetrads of points drawn from a real data set in two dimensions.

The domain $\mathscr{D}$ is now a projective 2-sphere (Fig. 7). We can throw half of this away and map the resulting hemisphere with antipodally collapsed boundary onto a cylindrical sleeve while preserving both independence and uniformity (Fig. 8). As we do not here care about the labellings of the points, the projective space can be decomposed into cells that are for us equivalent, and so we can work in any convenient one of these—here shown in the sleeve projection as the large spherical triangle (Fig. 9). The smaller spherical triangle (with a broken line boundary) will play an important role to be described in a moment.

At the boundary of the large spherical triangle at least two of the points of the tetrad coincide, while at its vertices there will be multiple coincidences. Such special cases (and near approximations to them) hardly represent what we think of as collinearities, and so we need to look at the central region of the large spherical triangle to find the true collinearities, if there are any.

Now the ancillary statistic M (position within the spherical triangle) determines the best fitting collinear tetrad, while the test statistic L tells us how far from collinearity the original tetrad is. Thus it becomes sensible to plot for each candidate tetrad, at the location specified by M, a marker symbol indicating a quantized value of L. In this example a shaded square indicates a very small value for L, while a

cross indicates an L-value up to 100 times greater. We then obtain *a single diagram containing all the relevant information* presented in a readily assimilable form. Of course we have taken the opportunity to clear away (by not plotting them if L exceeds the cut-off) those tetrads that are nowhere near to collinearity.

You will see that in this data set of 52 points (Broadbent's data) there are just four tetrads that are collinear to the standard currently being set (indicated in the diagram by four shaded squares). In fact the associated histogram for the L-values makes it plain that this number is not significantly large in relation to the numbers in the 100 nearby histogram cells.

We have not exhausted the geometric possibilities, however. There is further valuable information contained in the empirical distribution of the ancillary statistic M.

In the example shown you may have noticed that the spots marked by crosses are far from being uniformly distributed. Instead there is a tendency for them to be concentrated near the three corners of the large spherical triangle. This is a consequence of the fact that the empirical data set being used just happened to consist of two fairly distinct patches. The diagram is signalling that important fact to us in case we have not already noticed it.

Actually we are seriously concerned only with the shape points that fall in the smaller spherical triangle bounded by a broken line. This spherical triangle is defined by the requirement that each shape point *therein* represents an (ordered) linear tetrad ABCD satisfying an inequality of the form

$$\min(AB, BC, CD) \text{ is bounded below by } AD/5.$$

The choice of 5 as the divisor is somewhat arbitrary, but it seems reasonable. The inequality is to ensure that we do not overestimate the significance of a tetrad merely because it consists of three nearly collinear points and a fourth point close to one of these.

In fact the four very nearly collinear tetrads that we noticed in Fig. 9 all lie *outside* the central region delimited by the broken line (though one of them could be described as a 'near miss'). So that is another reason for not taking them too seriously.

Another attractive possibility would be to plot the data on an (L, R)-diagram, where here R is a uniformly distributed increasing function of the shortest *geodesic* distance on $\mathscr{D}$ from M to the boundary of the large spherical triangle, so that then the best candidates for serious near collinearity would have a small L and a large R. This last procedure will work for every value of k.

When such geometric viewing of the data lends support to the possibility that there may be causal factors generating close collinearities, then we can proceed to a numerical statistical analysis using the data-analytic techniques introduced by Wilfrid Kendall and myself in response to Broadbent's paper (Kendall and Kendall, 1980).

I hope that I have persuaded you how attractive this technique can be, if we push the algebra and analysis into the background and focus on the *geometry*.

It gives me much pleasure to propose the vote of thanks to my friend Colin Goodall. To prospective readers of the literature I add a word of caution: Professor Goodall's notation is different from that used in the series of theoretical papers on these topics dating from 1976, so care is necessary to avoid confusion.

John T. Kent (University of Leeds): Procrustes analysis forms the basis of an elegant and unified technique for the analysis of shapes from landmark data. However, many complications and variations arise in the presentation of the theory. Therefore it is useful to stand back for a moment and to ask what are the key ideas behind the application of Procrustes analysis in practice. I shall address these issues in the simple setting of defining an 'average shape' for collection of configurations in $K=2$ dimensions.

First I shall describe an alternative way to calculate a Procrustes average by using complex numbers. Represent a centred and scaled configuration of N landmarks by a complex vector $\mathbf{z}=(z_1, \ldots, z_N)^{\mathrm{T}}$, satisfying $\Sigma z_i=0$, and $\Sigma|z_i|^2=1$. Remember that $e^{i\theta}\mathbf{z}$, the rotation of $\mathbf{z}$ in the complex plane by an angle θ, and $\mathbf{z}$ represent the same shape. Next consider a collection of configurations $\{\mathbf{z}_j: j=1, \ldots, L\}$ in $\mathbf{C}^N$. A statistic which is invariant to rotations in the complex plane is the $N\times N$ complex sum of squares and products matrix, $\Sigma\mathbf{z}_j\mathbf{z}_j^*=T$, say. After some simple algebra, it turns out that the Procrustes average in Goodall's equation (2.5) is equivalent to the dominant eigenvector of T. There are links with a complex version of the Bingham distribution from directional data analysis and with principal component analysis. Unfortunately, this approach does not seem to generalize to $K\geqslant 3$ dimensions.

Next I would like to comment on the various ways in which an average shape might be defined. For example, we might use Goodall's equation (2.4) or (2.5) or Bookstein's (1986) co-ordinates, or the marginal

likelihood approach of Mardia and Dryden (1989a, b). Now in many applications shape data tend to be highly concentrated. After a little analysis, it turns out that all the methods listed here are asymptotically equivalent to one another to first order under high concentration. Further, the introduction of a more general covariance structure in any of the above approaches will make no difference, to first order. The reason that the covariance structure makes no difference is clear from the multivariate normal distribution, where the estimate of the population mean is the sample mean, regardless of the value of the covariance matrix. Therefore, it is important to emphasize the similarities between methods for constructing shape averages rather than to focus solely on their differences.

Lastly I would like to address the wider question of modelling shape by landmark methods. In many applications we start not with a set of landmarks but with a continuous outline of an object in the plane. This outline can be represented as a smooth function $f(t)$: $[0, 1] \rightarrow \mathbf{R}^2$. Here t might represent arc length, but other parameterizations may also be useful. From this functional point of view a natural comparison between two curves f and g can be defined in terms of the integrated squared difference

$$\int \| f^{(m)}(t) - g^{(m)}(t) \|^2 \, dt$$

between the function values themselves ($m = 0$), or their mth-order derivatives. f and g need to be registered with respect to one another, and care is needed in matching the meaning of t between f and g.

Unweighted Procrustes analysis can be viewed as a discrete approximation to this integral with $m = 0$. However, a larger value of m may yield a measure more sensitive to small bumps etc. in the outline. Some related ideas on the use of Markov models for outline data can be found in the work of Chow *et al.* (1988) and in Dryden and Mardia (1991). Further, the choice $m = 2$ is related to curvature, which I believe my colleague Professor Mardia will discuss later.

I am not entirely happy with the use of unweighted Procrustes analysis ($m = 0$) for the analysis of outline data because it takes no account of the fact that consecutive landmarks will be near one another in the outline. The use of derivatives overcomes this objection, but more work is needed to see how feasible the ideas will be in practice. Presumably this 'ordering' information can be incorporated in a weighted Procrustes analysis by a suitable choice of Σ_N. A rather different approach to take into account the layout of the landmarks in the analysis of shape change was given by Bookstein's (1986) biorthogonal grids in which shape change was modelled by a deformation of $\mathbf{R}^2$.

As you can see I have found the paper very stimulating and exciting. It gives me great pleasure to second the vote of thanks.

The vote of thanks was passed by acclamation.

Kanti Mardia (University of Leeds): I find the paper quite an important contribution to shape analysis. First, I would like to comment on the assumptions made in the paper. It is assumed that the number of landmarks and their locations $\{\mathbf{x}_t\}$, $t = 1, \ldots, N$, are known *a priori*. This may not always be true in practice and Mardia (1990) describes a curvature function method (for shape outlines) to obtain landmarks based on work with Ian Dryden. The factored covariance matrix may not be applicable in general; see Dryden and Mardia (1991) for a practical example. Also, the covariance matrix, at least for large N, should incorporate the neighbourhood structure of landmarks. For $K = 2$, a suitable model with population means $\boldsymbol{\mu}_t$ is

$$\{\mathbf{x}_t - \boldsymbol{\mu}_t\} = \text{circular bivariate AR}(p).$$

(In particular, the covariance matrix can be factored.) The advantage of such a model is that realistic shapes can be simulated easily. For deformable templates one models derivatives $\mathbf{x}_t - \mathbf{x}_{t-1}$ in this fashion (Chow *et al.*, 1988; Gough and Mardia, 1990); our approach differs from Chow *et al.* (1988) on various points of detail.

Contrary to the author's comments in Section 5, as far as I can see, the full and marginal procedures are based on the same assumptions. In particular, the marginal approach does not require a constant co-ordinate system. I agree that the computational burden increases rapidly with N for the marginal approach. However, convenient co-ordinate systems exist on shape space, e.g. $\mathbf{W}$ in Goodall and Mardia (1990a), which facilitate the direct examination of data and calculation of summary statistics such as the mean vector and covariance matrix. Also, for cluster data in any dimension, the marginal likelihood estimators will be approximately the same as the sample mean vector and covariance matrix, provided that we keep away from the singularities of the co-ordinate system.

In on-line gesture recognition, one wants to discriminate between hand-drawn shapes by using a mouse on a computer. Here, the shapes are identical under non-linear scaling, e.g. for the operation 'transpose' the following objects represent the same command:

Thus, the work involves non-linear scaling and we require an extension to the existing shape theory. For the above shape, the obvious landmarks are corners and appropriate angles can be shape variables but this is not true for other shapes (Mardia *et al.*, 1990).

I look forward to seeing applications of Procrustes and other techniques in three dimensions, e.g. object recognition in computer vision.

W. S. Kendall (University of Warwick, Coventry) **and H. L. Le** (University of Cambridge): We congratulate the author on a paper which we found very stimulating.

It would be useful to summarize differences between shapes of labelled k-ads in terms of differences between shapes of triads obtained from the k-ads via (appropriately restricted) affine combinations of vertices selected in some optimal fashion. This would make available the spherical blackboard of D. G. Kendall for display of 'the most prominent change of shape'; moreover the affine combinations forming the representative triads could be displayed as weightings of the landmarks making up the original k-ads. We have begun to think about representations which might be suitable in terms of being economical, visualizable and interpretable.

So far we have made only small progress with this idea. Our starting point was a fact pointed out by Goodall to one of us: consider the geodesic defined by the change of shape from one labelled k-ad to another, and project down to the sphere Σ_2^3 by selecting three combinations of points. Except in degenerate cases, the *locus* of the projection (but not the projected curve itself) forms an arc of a small circle on Σ_2^3.

For data analysis we must make an optimal selection. Options include making the locus as great a circle as possible, or maximizing the resulting change of triad shape. But (if $k \geqslant 4$) there are selections for which the triad remains of constant shape and so the first option yields a poorly posed problem. The second option is a little more promising (especially since the interpolating small circle can be added to the resulting diagram). It is related to an optimizing problem, reminiscent of both Procrustes and canonical correlation analysis, which we are investigating.

Finally, we comment on three points of detail.

(a) W. S. Kendall's use of the singular value decomposition actually arises in the *sequel* (Kendall, 1990a) to the paper referred to by Goodall.
(b) The *mean shape* is an instance of a generalized notion of expectation due to Fréchet (and previously E. Cartan), which has recently been applied to the probabilistic theory and non-linear partial differential equations (Kendall, 1990b).
(c) There are yet more general notions of shape. Ambartzumian (1990) introduces the *affine shape* of (for example) a tetrad of points in the plane.

J. C. Gower (University of Leiden): Everyone is agreed that similar figures have the same shape and I welcome Professor Goodall's paper that provides an inferential structure for testing similarity.

I shall say something about alternative geometries for representing shape change and something about the measurement of shape itself. The co-ordinate matrices $\mathbf{X}_i$ generate sets $\mathscr{E}$, $\mathscr{D}$ and $\mathscr{B}$ whose members are matrices of distances $\mathbf{E}_i$, of square distances $\mathbf{D}_i$ and of inner products $\mathbf{B}_i$. Matrices $\mathbf{E}$, $\mathbf{D}$ and $\mathbf{B}$, corresponding to the Procrustes group average $\overline{\mathbf{X}}$, belong to these sets. All sets are invariant to orthogonal transformations and $\mathscr{D}$ and $\mathscr{E}$ are also invariant to translations. $\mathscr{E}$ is the natural space for studying distributional effects when lengths are measured rather than landmark co-ordinates but, unlike $\mathscr{D}$ and $\mathscr{B}$, is not convex. Take the subdiagonal elements of any of these matrices as co-ordinates of points in at most $n(n-1)/2$ dimensions and labelled E_i, D_i, B_i $(i=1, \ldots, L)$. The average matrices $\overline{\mathbf{E}}$, $\overline{\mathbf{D}}$ and $\overline{\mathbf{B}}$ form useful summaries and a conical projection of $\overline{D}$ onto an edge of $\mathscr{D}$ gives the group average of individual differences scaling (Carroll and Chang, 1970). $L \times L$ matrices formed from the distances between pairs of points in $\mathscr{D}$ or $\mathscr{E}$, or from pairwise Procrustes statistics (Gower, 1970), may be displayed by multidimensional scaling methods to reveal patterns among the L configurations (e.g. clusters or trends); this approach might be useful with the rat data. The points can be constrained to

represent matrices of constant size by scaling $\mathbf{1}'\mathbf{D}_i\mathbf{1}=n$, so that the B_i and D_i lie on hyperplanes orthogonal to $\mathbf{1}_c$ and the E_i lie on a hypersphere; normally E and D do not lie within these constrained spaces.

The generalized Procrustes analysis statistic m^2 may be written geometrically,

$$nm^2=\Sigma E_i\bar{E}^2+(O\bar{E}^2-OE^2),$$

and algebraically,

$$nm^2=\Sigma(\mathbf{1}'\mathbf{D}_i\mathbf{1})/L-\mathbf{1}'\mathbf{D}\mathbf{1}=\mathbf{1}'(\bar{\mathbf{D}}-\mathbf{D})\mathbf{1}.$$

In general $\bar{E}$ and E do not coincide and $\bar{\mathbf{E}}$ may not even represent a distance matrix; nevertheless the result suggests that $\mathbf{E}$ is close to $\bar{\mathbf{E}}$. $\mathscr{D}$ is convex so certainly contains $\bar{D}$, and m^2 is the projection of $\bar{D}-D$ onto the central ray $\mathbf{1}_c$. When the $\mathbf{X}_i$ are centred at their centroids, $\mathscr{B}$ is polar to $\mathscr{D}$ (Critchley, 1986) and, because $\mathbf{1}'\mathbf{B}_i\mathbf{1}=-\mathbf{1}'\mathbf{D}_i\mathbf{1}/2n$, similar geometries hold in $\mathscr{B}$ and $\mathscr{D}$. Distances in the two spaces are linked by the expression

$$\Sigma(d_{jk}^{(1)2}-d_{jk}^{(2)2})^2=4\Sigma(b_{jk}^{(1)2}-b_{jk}^{(2)})^2+(n+2)\Sigma a_j^2+(\Sigma a_j)^2$$

where $a_j=b_{jj}^{(1)}-b_{jj}^{(2)}$. With standardization, $\Sigma b_{jj}^{(1)}=\Sigma b_{jj}^{(2)}$, so that $\Sigma a_j=0$. If further $b_{jj}^{(1)}=b_{jj}^{(2)}$, then distances in the two spaces are equivalent. The b_{jj} give the squared distances of the landmark points from their centroid and may be isolated and analysed as measures of what might be termed *size-shape*. Then $\beta_i^{-1/2}\mathbf{B}_i\beta_i^{-1/2}$ with $\beta_i=\text{diag}(\mathbf{B}_i)$ represents purely *angular-shape*. Non-centroid origins, such as the generalized circumcentre (Gower, 1985) or the mediancentre, define useful B-matrices.

However, shape differences are often localized, so global measures of shape are suspect.

Ian L. Dryden (University of Leeds): Procrustes techniques in Euclidean space suggest analogous models in the shape space Σ_K^N, and I shall concentrate on the important $K=2$ case. Consider the following densities, with respect to Kendall's (1984) uniform measure in Σ_2^N:

$$a_N(\varkappa)\exp\{\varkappa\cos(2\alpha)\}, \tag{1}$$

$$b_N(\varkappa)\exp(4\varkappa\cos\alpha), \tag{2}$$

$${}_1F_1[2-N;\ 1;\ -\varkappa\{1+\cos(2\alpha)\}]\exp\{\varkappa\cos(2\alpha)-\varkappa\}, \tag{3}$$

where $0\leqslant\alpha<\pi/2$ is the Riemannian distance from the population shape; $\varkappa\geqslant 0$ is a concentration parameter, ${}_1F_1(c;\ d;\ x)$ is the confluent hypergeometric function (a terminating series above) and

$$a_N^{-1}(\varkappa)=\exp(-\varkappa)\,{}_1F_1(1;\ p-1;\ 2\varkappa),$$
$$b_N^{-1}(\varkappa)=1+[\sqrt{\pi}\,(2\varkappa)^{5/2-N}(N-2)!\{I_{N-3/2}(4\varkappa)+\mathbf{L}_{N-3/2}(4\varkappa)\}],$$

where $I_\nu(\,)$ and $\mathbf{L}_\nu(\,)$ are the modified Bessel (first kind) and modified Struve functions. If $N=3$ then $a_3=\varkappa/\sinh\varkappa$ and $b_3=8\varkappa^2/\{1-\exp(4\varkappa)+4\varkappa\exp(4\varkappa)\}$.

The maximum likelihood estimate (MLE) of shape using expression (1) is the same as the Procrustes-with-scaling mean shape defined in equation (2.5) of the paper. For $N=3$ expression (1) is the density of the Fisher distribution on $\Sigma_2^3\cong S^2(\frac{1}{2})$, used by Mardia (1989a). The MLE of shape using expression (2) is the same as the Procrustes-without-scaling mean shape defined in equation (2.4). Thus, the two Procrustes approaches can be compared with the marginal likelihood approach (Mardia and Dryden, 1989a), where the shape density under the isotropic Gaussian model is expression (3).

All three distributions are asymptotically normal as $\varkappa\to\infty$ and uniform if $\varkappa=0$. The three densities are very similar for much of the range and in practice the models give similar shape estimates. However, expression (2) has particularly light tails and so the Procrustes-without-scaling estimate will be the most affected by outliers.

Professor Goodall and I have empirically compared the Procrustes and marginal MLE shape estimates for some moderately dispersed data (p. 300). In Fig. 10 the shapes of triangles from 20 neural spines of T2 mouse vertebrae are given, together with various shape estimates, in Kendall's shape co-ordinates $\{\text{Re}(u^*),\ \text{Im}(u^*)\}$. For these data the Procrustes-with-scaling estimate (0.0934, 0.9799) and marginal shape estimate (0.0885, 0.9838) are almost identical, the Procrustes-without-scaling estimate (0.0653, 1.0017) is similar but Bookstein's (1986) arithmetic mean $(-0.1086,\ 0.8315)$ is appreciably different. This example suggests that the difference between the so-called 'full' and marginal MLE is negligible for practical purposes.

The marginal MLE approach can also be extended to general covariances for $K=2$ dimensions (Dryden and Mardia, 1991). Like the Procrustes shape estimates for the rat data in Section 14.4, the marginal shape MLE is not greatly affected by the choice of covariance structure, if variations are small. However,

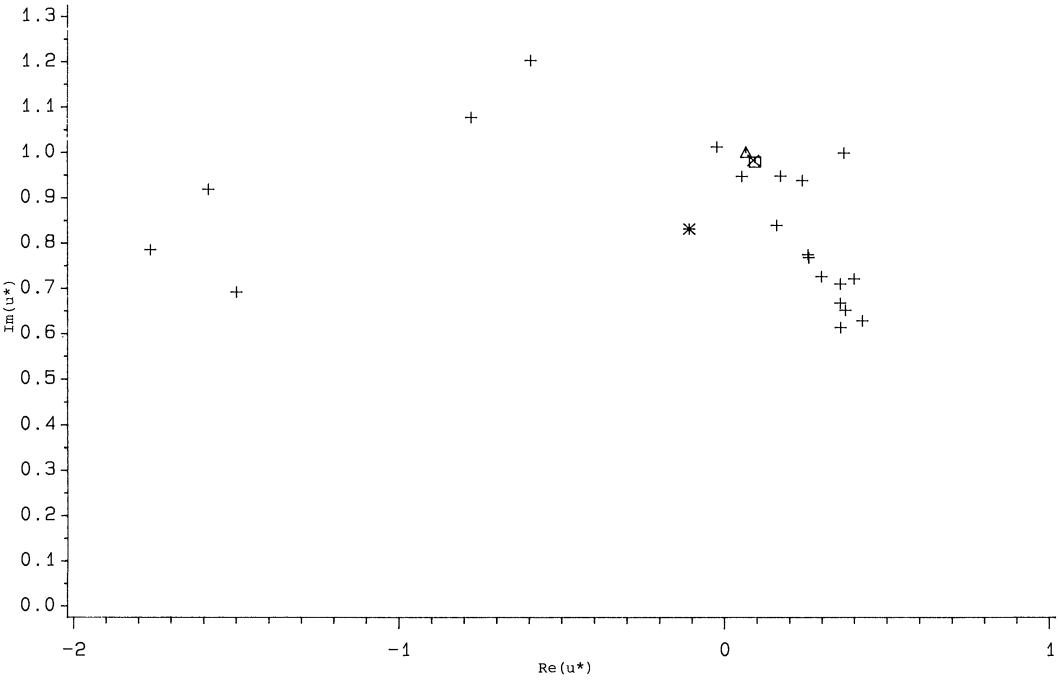

Fig. 10. Shapes of 20 T2 neural spine triangles (+) in Kendall's shape co-ordinates, with the Procrustes-with-scaling (□), Procrustes-without-scaling (△), marginal MLE (×) and Bookstein's mean (∗) shape estimates

care must be taken with inference unless the covariance structure is simple. The major problem is that singular landmark distributions can lead to non-degenerate shape distributions. Are the general Procrustes methods of Section 10 easy to work with?

Alistair Walder (University of Leeds): I would like to make a comment relating to Sections 5 and 6, extending the Gaussian and the two-sample model. With Professor Mardia I have been looking at the joint shape density of two figures in $\mathbf{R}^2$, X and Y say, which are correlated over time. (For example, consider X-rays of the same patient taken at different occasions.)

Our model is simple:

$$X = \mu_X + E_X,$$
$$Y = \mu_Y + E_Y;$$

cf. equations (6.1) and (6.2); our approach is 'marginal' in the sense of Section 5. No attempt is made to estimate the translation, rotation or scale. However, the model for X and Y also differs in that E_X and E_Y are not independent. Let $\mathbf{e}$ be the $4N \times 1$ vector obtained by unstacking the columns of E_X, then E_Y (both $N \times 2$). We assume that $\mathbf{e} \sim N(0, \Sigma)$ and that Σ can be factorized, as in Mardia (1984) (where the sites are now landmarks) or this paper, as

$$\begin{pmatrix} \sigma_1^2 & 0 & \rho\sigma_1\sigma_2\cos\psi & \rho\sigma_1\sigma_2\sin\psi \\ 0 & \sigma_1^2 & -\rho\sigma_1\sigma_2\sin\psi & \rho\sigma_1\sigma_2\cos\psi \\ \rho\sigma_1\sigma_2\cos\psi & -\rho\sigma_1\sigma_2\sin\psi & \sigma_2^2 & 0 \\ \rho\sigma_1\sigma_2\sin\psi & \rho\sigma_1\sigma_2\cos\psi & 0 & \sigma_2^2 \end{pmatrix} \otimes I_N.$$

This includes the minimal assumptions necessary for a sensible model. We assume independence between landmarks within figures, corresponding landmarks are correlated and perturbed similarly over time and the angle ψ accommodates the rotation of base edges involved in edge superimposition.

Despite these simplifying assumptions the resulting distribution is not aesthetically pleasing but can be written in closed form. It is then not surprising that our marginal maximum likelihood estimates have to be estimated numerically.

I would like to finish by asking Professor Goodall whether Procrustes methods can be easily extended to correlated figures.

Toby Lewis (University of East Anglia, Norwich): I wonder whether Professor Goodall has considered a shape problem of prime practical interest to palaeomagnetists, the analytic comparison of apparent polar wander paths (APWPs). This falls between the landmark situation and the situation discussed by Professor Kent and Professor Mardia where there are curves or outlines.

Suppose that we take some rock specimens, say in Australia, and obtain from each specimen an estimate of the position of the palaeomagnetic pole as it has moved around over hundreds of millions of years. If the ages of the specimens can either be estimated or at least ranked, the positions can be represented by a time-ordered sequence of points on the surface of a unit sphere, forming a path—an APWP. Fig. 11 shows an APWP for Australia (adapted from Embleton *et al.* (1983)). If the same is done in Africa or North America, the same path should be obtained, apart from noise. One reason why it is not is because the continents have moved relative to one another over geological time. What rotation will bring the two paths into best alignment (whatever 'best' may mean), and after this rotation are the paths reasonably consistent (on some criterion)? Fig. 12 (same source as Fig. 11) shows an APWP

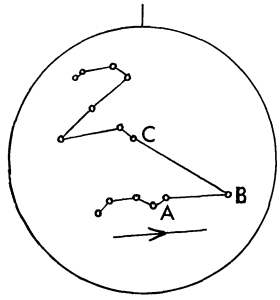

Fig. 11. Data set for Australia

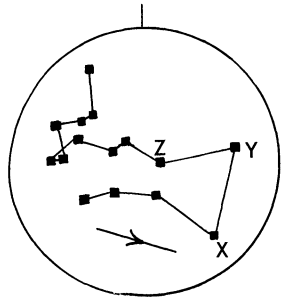

Fig. 12. Data set for North America

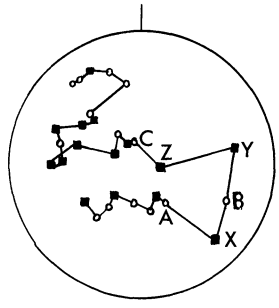

Fig. 13. Pooled data set

for North America after rotation to match the data in Fig. 11. The statistical problem takes various forms according to the information available. In the case illustrated here, only the separate chronological orders for the two sets of pole positions are known, and not the actual ages or even the pooled chronological order. The data points in Figs 11 and 12 are not landmarks because they do not correspond in pairs; in fact, there are unequal numbers of points in the two data sets. Fig. 13 (same source) shows an APWP based on the combined information; path ABC in Fig. 11, for example, is replaced by path AXBYZC in Fig. 13 because there were no Australian observations for the times covered by North American data points X, Y, Z.

There are many such problems in APWP analysis. I would like to ask Professor Goodall whether he has already dealt with this kind of problem and what his words of wisdom about it are.

D. Stoyan (Bergakademie Freiberg) **and H. Ziezold** (Gesamthochschule Kassel): We congratulate Professor Goodall on his excellent paper. Also we believe that the use of Procrustes methods is indeed of very great value for shape and form statistics. We have used Procrustes distances as a starting point for cluster analysis (Stoyan, 1990a) and multivariate scaling.

In particular we are impressed by the theory of Gaussian landmark models with non-independent and non-spherical displacements. In own studies of shapes of hands we found that the original Bookstein model (with independent and identically distributed (IID) spherical displacements) is quite unrealistic for modelling the intrapopulation variability (Stoyan, 1990b). This paper shows that statistical methods for the χ-distribution permit the direct estimation of the distances of landmark centres and the corresponding variances. The more general models suggested in Section 5 look very promising.

Are there biological objects for which the original Bookstein model is a good model for the intrapopulation variability and not for measurement errors only?

The determination of mean shapes and forms is a very important problem for practical data analysis of shapes and forms. These Procrustes means are merely examples of Fréchet's definition of expected elements of random elements in metric spaces (Ziezold, 1977). Ziezold has proved a strong law of large numbers for IID elements in a separable metric space thus giving a deeper mathematical justification for these means. In the same paper he applied the theory to samples of forms and showed some elementary properties. Furthermore, he analysed Gower's algorithm for simulated samples experimentally (Ziezold, 1989). Like Professor Goodall, he found that in practical important cases the algorithm needs only 3–5 iterations. But if, for example, the sample consists of many figures which are generated by three points independently chosen by the uniform distribution in the unit circle, the limit of the then very slowly convergent algorithm also depends on the starting figure. The main limits are

(a) a triangle with two points near together and
(b) a nearly collinear triangle with one vertex near the arithmetic mean of the two other vertices.

As far as we know, general mathematical criteria for the convergence or even for the speed of the convergence of the algorithm have not yet been found.

We should also like to know the behaviour of the Procrustes mean shapes or forms for $L \to \infty$ in the Bookstein model. In general, these means are not unbiased estimators of the shape or form generated by the expected points.

Trevor Hastie and Eyal Kishon (AT&T Bell Laboratories, Murray Hill): With M. Clark of Monash University, Melbourne, we have been developing a statistical model for signatures. As part of the definition

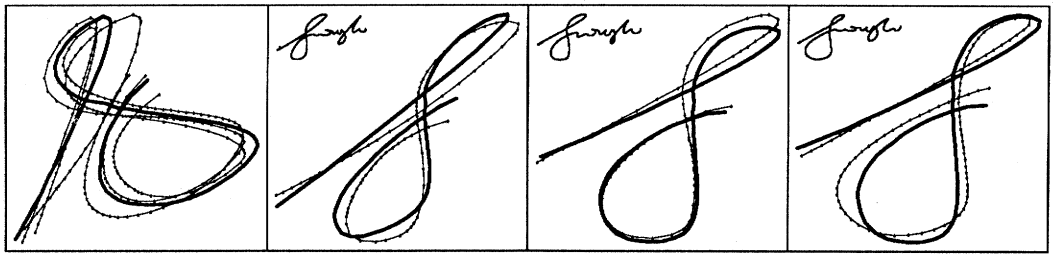

Fig. 14. (a) Mean $\hat{S}$ (bold) corresponding to $L = 3$ renditions of Suresh's signature with $N = 100$, together with the three projections $Y_i \hat{A}_i = Q_i Q_i^{\mathrm{T}} \hat{S}$; (b)–(d) individual signatures, together with their modelled versions $\hat{Y}_i = \hat{S} A_i^{-1}$

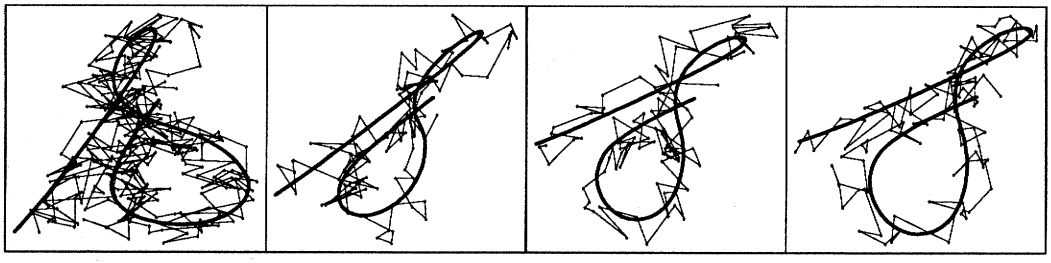

Fig. 15. (a) Smooth mean signature (bold) with the projections of the three contaminated signatures superimposed; (b)–(d) individual contaminated signatures, together with their smooth modelled versions

of the *mean signature* of an individual, we (independently) arrived at an expression similar to expression (2.1). Here we outline a simplification achieved by relaxing the assumptions underlying expression (2.1), as well as a 'smooth' version thereof (Figs 14 and 15).

Suresh signs his name L times on a pressure pad, which samples the co-ordinates and pressure of the pen 1000 times a second. Assume that we have (centred) signature matrices Y_i, $i=1, \ldots, L$, each of dimension $N\times 2$ and representing N *corresponding* points from the leading S of the signatures. We refer to these subsampled Y_i as the 'signatures'.

Similar to Gower's (1975) definition (2.1) and (2.5) in the paper, we define the average (preshape) signature $\hat{S}$ to be the minimizer of

$$\sum_{i=1}^{L} \| Y_i A_i - S \|^2 \tag{4}$$

over $S_{N\times 2}$ and A_i but allow the A_i to be general 2×2 matrices. The A_i allow for the deformations that Suresh might make to his S in different renditions of his signature, including shearing. To avoid degeneracies we impose the constraint $S^{\mathrm{T}}S=I$ (for convenience).

Problem (4) has a simple solution that does not require iteration. If $Y_i=Q_iR_i$ is the QR decomposition of the ith signature, then the optimal A_i is $\hat{A}_i=R_i^{-1}Q_i^{\mathrm{T}}S$ and hence $Y_i\hat{A}_i=Q_iQ_i^{\mathrm{T}}S$. So at the minimum expression (4) is

$$\sum_{i=1}^{L} \| (Q_iQ_i^{\mathrm{T}}-I)S \|^2 = \sum_{i=1}^{L} \mathrm{tr}(S^{\mathrm{T}}M_iS)$$

$$= L\,\mathrm{tr}(S^{\mathrm{T}}\overline{M}S)$$

where $\overline{M}$ is the average of the residual projection operators $M_i=(I-Q_iQ_i^{\mathrm{T}})$. Since each of the M_i are symmetric and non-negative, so is $\overline{M}$. Minimizing $\mathrm{tr}(S^{\mathrm{T}}\overline{M}S)$ subject to $S^{\mathrm{T}}S=I$ is a well-known eigenvector problem, with solution $\hat{S}$ a basis for the eigenspace corresponding to the two smallest eigenvalues of $\overline{M}$.

Typically the rows of the Y_i are sampled at equal arc lengths along the signature, which is a smooth curve, and so they define a smooth (discrete) curve parameterized by the index set $1, \ldots, N$. Usually this results in a similarly smooth mean $\hat{S}$. The Y_i themselves may be smooth curves contaminated with noise. These and other considerations motivate us to consider a smooth generalization of problems (2.1) and (2.5), and our problem (4). S will be a smooth curve if each of its columns is smooth. We can achieve this by augmenting problem (4) with a quadratic *roughness penalty* of the form $\lambda\,\mathrm{tr}(S^{\mathrm{T}}\Omega S)$; one candidate for $\Omega_{N\times N}$ is the integrated second squared derivative matrix corresponding to a smoothing spline (as in Rice and Silverman (1991)). In our case the solution for S is then a basis for the eigenspace corresponding to the two smallest eigenvalues of $L\overline{M}+\lambda\Omega$. Fig. 15 shows the same signatures as in Fig. 14, but contaminated with noise.

Christopher G. Small (University of Waterloo): I would like to pick up the point from Section 16 that the exact shape densities are 'heavy going'. This is undoubtedly true, but many avenues remain open. For example we could use Laplace approximations to calculate shape densities for quite general models. Consider a set $\mathbf{x}=(\mathbf{x}_1, \ldots, \mathbf{x}_n)$ of independent planar landmarks with the ith landmark $\mathbf{x}_i$ centred at $\boldsymbol{\mu}_i$. We suppose that the landmarks have common error distribution with density f around

their respective centres $\boldsymbol{\mu}_i$. In Bookstein shape co-ordinates $\mathbf{z}=(\mathbf{z}_1, \ldots, \mathbf{z}_{n-2})$, the induced shape density $\tilde{f}(\mathbf{z})$ is a twofold integral over complex co-ordinates. Let $\hat{\mathbf{x}}=(\hat{\mathbf{u}}, \hat{\mathbf{u}}+\hat{\mathbf{v}}, \hat{\mathbf{u}}+\hat{\mathbf{v}}\mathbf{z}_1, \ldots, \hat{\mathbf{u}}+\hat{\mathbf{v}}\mathbf{z}_{n-2})$ be a translated, rescaled and rotated version of $\mathbf{x}$ such that $(\hat{\mathbf{u}}, \hat{\mathbf{v}})$ maximizes the value of

$$g(\mathbf{u}, \mathbf{v})=\|\mathbf{v}\|^{2n-4} f(\mathbf{u}-\boldsymbol{\mu}_1) f(\mathbf{u}+\mathbf{v}-\boldsymbol{\mu}_2) \prod_{i=1}^{n-2} f(\mathbf{u}+\mathbf{v}\mathbf{z}_i-\boldsymbol{\mu}_{i+2}).$$

Then the Laplace approximation to the shape density is given by

$$\tilde{f}(\mathbf{z}) \sim 4\pi^2 |A(\hat{\mathbf{u}}, \hat{\mathbf{v}})|^{-1/2} \|\hat{\mathbf{v}}\|^{2n-4} f(\hat{\mathbf{u}}-\boldsymbol{\mu}_1) f(\hat{\mathbf{u}}+\hat{\mathbf{v}}-\boldsymbol{\mu}_2) \prod_{i=1}^{n-2} f(\hat{\mathbf{u}}+\hat{\mathbf{v}}\mathbf{z}_i-\boldsymbol{\mu}_{i+2})$$

where A is the 4×4 Hessian matrix of $\log g$ with respect to the co-ordinates $\mathbf{u}$, $\mathbf{v}$. The computation of the determinant of A is tedious but can be done with symbolic packages such as MAPLE. This task can be simplified if the $(\mathbf{u}, \mathbf{v})$-co-ordinate system can be reparameterized to give orthogonal co-ordinates.

Now suppose that we let the vector of means $\boldsymbol{\mu}=(\boldsymbol{\mu}_1, \ldots, \boldsymbol{\mu}_n)$ be an unknown parameter. We might suppose that $\boldsymbol{\mu}$ lies in some constrained family of configurations, and that the intention is to match the shape of the data to some lower dimensional constrained family of shapes (much as in regression where one constrains the estimated response to fall on a straight line). Maximum likelihood can be used and the approximate density can be maximized over the constrained space of values of $\boldsymbol{\mu}$. As the Laplace approximation involves the maximization of g in $\mathbf{x}$ it is natural to maximize jointly in $(\mathbf{x}, \boldsymbol{\mu})$ rather than solving for $\hat{\mathbf{x}}$ first. The result is an extension of Procrustes techniques to models with a general error structure. In principle, the Laplace approximation can be accomplished for data in any dimension. However, for landmarks in dimensions higher than 2 we lose the advantages of complex arithmetic which make the formulae simple.

The possibility of approximating shape distributions should not deter us from calculating these distributions exactly whenever possible. In some cases the integrals can be computed exactly. Among such examples the distribution of shape for the vertices of Delaunay simplexes of a Poisson process is particularly elegant.

The future of shape theory will inevitably require much more sophisticated models than those in which landmarks have independent errors. I am reminded that statistics has been largely concerned with the estimation of location and dispersion for much of its history. However, these parameters are finite dimensional. Everything that remains is shape.

The following contributions were received in writing after the meeting.

John Bacon-Shone (University of Hong Kong): Professor Goodall's paper is very valuable in that it sets out a general structure for the analysis of shape problems. However, I believe that there are important classes of shape problems, for which his approach, as it stands, may not be the most appropriate. For example, in his rat skull example, there is information in the ordering of the landmarks. It is possible that the major point of interest is the interpoint distances of successive landmarks (or some other subset of the interpoint distances). In this case the most appropriate strategy would be to use compositional data analysis (Aitchison, 1986) to look at the interpoint distances relative to the sum of the distances. Conversely, it could be argued that what is of interest may not be the interpoint distances, but rather their complement in shape space, i.e. the variation in orientation of the landmarks, once variation in successive interpoint distances is taken into account. This can be done by regressing the standardized landmarks on the vector of log-ratios of interpoint distances. A nice example of when this idea may be of use is in assessing the gait of children via landmark data on joints taken over time for many different children. In this case, it is of interest to know the extent to which body orientation is determined by bone lengths. I am currently investigating this together with a physiotherapist.

Fred L. Bookstein (University of Michigan, Ann Arbor): In morphometrics, the interpretation of sample variation in shape space is inseparable from the geometry of the mean shape. Morphometric explanations are governed by two descriptive languages at once. One description deals with global aspects of change, measured via first derivatives, the other, with local aspects (contrasts between patterns in groups of landmarks at greater or lesser separation), measured via second derivatives. The affine subspace Goodall mentions supports a global (first-derivative) metric, e.g. anisotropy, the complementary subspace, an incommensurate second-derivative metric (e.g. some analogue of 'bending energy'). This decomposition of directions in shape space depends *strongly* on the mean shape. But in Procrustes analysis the geometric

properties of shape space are almost independent of the mean form as long as we stay away from the collinearity set. Furthermore, statistical space is rotatable (up to a covariance structure) near the mean form, but in biological applications it is not meaningful to rotate between the two subspaces.

So for applications in morphometrics Procrustes analysis is not only scientifically inefficient (in discarding the crucial information about means) but also is mired in an inappropriate algebra of spaces and their descriptors. The metric that Goodall claims is 'natural' is not for several traditional biometric purposes. For example, the principal components of a sample in shape space must be taken in two sets: one pair within the affine subspace, and another set, in the complementary space, that emerges in order of geometrical scale. Analysis of data for growing rats indicates considerable independence of trends and regulatory precision in these two subspaces over normal growth. To understand the processes by which they grow, then, the descriptions, and hence the metrics underlying, must be kept separate. Similarly, the idea of any unitary regression in shape space presumes one single metric in which the residuals should 'naturally' be minimized. There being no such metric, regressions must be carried out in both of these subspaces, and the results then combined via a covariance structure.

Contrary to Goodall's suggestion, metrics for understanding the natural world arise only in nature, not in the mathematics of lower triangular matrices. Living organisms know, as the algebra of the QR decomposition does not, that adjacent landmarks covary for different reasons than landmarks at a distance. Therefore, although Goodall's method can be made to yield fully efficient tests of mean differences in designed experiments, I would not apply it in any more subtle sort of biometric investigation. It is not statistical science. The geometry of shape analysis for organisms is bound up with the details of the mean form in a manner for which the notation of Cartesian product covariance matrices in a single metric is remarkably ill suited.

Frank Critchley (University of Warwick, Coventry): I would like to comment on two points, and then to note some related work.

Distance geometry

I agree with Professor Gower that, among other possibilities, it is natural to consider appropriate invariants directly when analysing shape. In particular, the $N \times N$ matrix of squared Euclidean interpoint distances D takes out location and orthogonal transformation while, trivially, multiplication by the appropriate constant takes out scale. The geometry of the closed convex cone $\mathbf{D}$ of all such matrices then comes to the fore. In certain respects this geometry is pleasingly simple. The question, then, is: which geometry is better for which statistical analysis?

Degrees of freedom

The many-faceted nature of shape is reflected in the large number of degrees of freedom in the test statistics described in Section 14. Two obvious and complementary questions arise.

(a) Can the overall test statistics be decomposed into preferably orthogonal parts, each having a clear interpretation?
(b) Against *which* more specific alternative hypotheses does the test of equal shape have greatest power?

An orthogonal decomposition introduced in Critchley (1988) may be of value when question (a) is considered in the context of the distance geometry above.

Critchley (1980, 1986, 1988, 1991) contain work relevant to distance geometry, some of which we briefly note here. The interior of $\mathbf{D}$ is precisely the set of all those D for which the corresponding configurations have maximal dimensionality $N-1$. Most of the interest of $\mathbf{D}$ is, then, in its boundary! Critchley (1991) obtains the support cone at each boundary point of $\mathbf{D}$ and shows that $\mathbf{D}$ admits a central ray which corresponds to the regular simplices. The angle which a matrix D makes with this central ray can be viewed as a continuous measure of the Euclidean dimensionality of the configurations associated with D, in a certain natural sense. The larger this angle, the lower the dimensionality is. Critchley (1986), proposition 12, gives the polar cones of $\mathbf{D}$ and of the set $\mathbf{B}$ of all centred inner product matrices. In each case the polar is taken with respect to the relevant linear hull of the original cone. Unsurprisingly, $\mathbf{B}^0 = -\mathbf{B}$. Much more interestingly, $\mathbf{D}^0 = \mathbf{B}^1$, where B^1 denotes B with all its diagonal entries set to 0. Critchley (1988) sets up the mathematical machinery to analyse how the properties of $\mathbf{D}$ are related to, and so can be obtained from, those of five other cones. Critchley (1980), p. 223, notes the tiresome non-convexity that arises when we *do not* square the Euclidean distances in D but that this problem does not arise when $N=3$, when the non-zero boundary of $\mathbf{D}$—or of its elementwise square root—consists of the collinear triangles.

Nicholas Lange (Brown University, Providence): I have a comment on an applied modelling, data analytic aspect of Professor Goodall's work, perhaps most relevant to Section 5.

When we observe a large number of short series of repeated shapes for each individual unit under study, it is quite natural to decompose total variation into its within- and between-individuals components of variation and covariation. Our joint work deals with the simplest interesting non-trivial case of longitudinal series of planar triangles whose trajectories are modelled on Kendall's sphere by using simple growth curves. Statistical summaries and inferences regarding changes in shape over time depend strongly on the choice of a Euclidean or non-Euclidean shape space in our context. Magnitudes of discordances between conclusions drawn using one space or another seem to depend directly on the magnitudes of total shape change throughout the observation period.

In addition to analysing differences in means, it is always important to ask 'What happened to the (co)variance?'. When we presume homogeneity of within-individual, landmark-specific disturbances, covariance component analyses for repeated shapes in a growth curve setting are direct extensions of somewhat classical univariate methodology to a multivariate context, with perhaps a change of space. Yet when common variance is an implausible assumption, as would be the case when different landmarks are identifiable with differing degrees of precision, for instance, obtaining maximum likelihood estimates of primary and nuisance parameters is more difficult.

Recent advances in methodology for hierarchical Bayesian computations with high dimensional models, with or without an assumption of conjugacy between likelihood and prior, make the restrictive common variance assumption unnecessary (see, for example, Lange *et al.* (1990)). We could use 'Gibbs sampling' technology in the statistical analysis of repeated shapes with different individual and landmark-specific disturbances in a straightforward albeit challenging and computationally demanding analysis.

Equally demanding would be to carry out the type of experimental design to obtain the data required for estimation of landmark-specific (co)variances reliably. To do so properly, we would require multiple sets of landmark co-ordinates for each individual sampling unit, one set from each of a moderate number of independent 'digitizers' (human or machine). Even if we were able to obtain such a nice data set, allowing and fitting heterogeneous (co)variances could perhaps not make all that much difference in practical conclusions when compared with applications of simpler approaches in some cases. Yet if the various digitizers are not human, but are instead a moderate number of competing medical imaging modalities, then perhaps more challenging and computationally intensive models and methods are required to derive the most information from available data.

Subhash Lele (Johns Hopkins University, Baltimore): I would like to thank Professor Goodall for asking me to comment on his mathematically impressive paper. The optimality properties of the Procrustes method of superimposition for estimating the mean form and shape are important to know. However, sometimes I wonder whether, in pursuit of mathematical pleasure, we statisticians are losing sight of the scientific problems that we purport to solve.

My main problem with superimposition methods (Lele, 1991) is that the definition of shape or form difference is strongly tied to the choice of the loss function used for superimposition. Following Goodall and Bose (1987), if we write $Y = b(X + J)B + t$ and call J the shape difference between X and Y, J is non-identifiable. This is why in Fig. 4 we obtain three different inferences about how the same two groups differ in shape. Scientifically this is unsettling. A biologist is usually not interested in merely testing whether the two populations are equal in shape but mainly in knowing where and by how much the two shapes or forms are different. Localization of form difference is problematic when using superimposition methods of form comparisons. I do not know how we can approach the study of allometry and other biological problems with a non-identifiable form and shape difference. We can circumvent this problem of non-identifiability by using the Euclidean distance matrix representation for the landmark data. See Lele and Richtsmeier (1991a, b) and Richtsmeier and Lele (1990).

One technical point: it is not clear to me why Σ_N is estimable. Is it not the same problem as discussed in Neyman and Scott (1948) where, if the X_i are independent $N(m + t_i, \sigma^2)$, σ^2 is not estimable unless there are at least two observations with the same t_i? We can take Σ_N to be diagonal which implies that landmarks are varying independently of each other. This seems to me a highly unrealistic assumption. For the general Σ_N, it is not clear how we can use the statistic suggested in Section 13, paragraph 1.

The **author** replied later, in writing, as follows.

I wish to thank the discussants for their informative, generous, and always helpful and interesting

comments. Dr Small notes that 'everything that remains is shape', and indeed the breadth of the discussion reflects that fact. In this rejoinder I shall consider first the more technical issues (algorithms and models) before turning to the discussion of shape differences.

Spectral methods

Professor Kent shows that the mean shape for planar figures is the eigenvector corresponding to the largest eigenvalue of an $N \times N$ complex matrix. Rohlf and Slice (1990) give a similar solution in the affine case, now cast into an elegant QR form by Dr Hastie and Dr Kishon. In general, if the two-figure superimposition is a linear regression (for the planar shape case, see the end of Section 7), then the mean of L figures is the solution of a spectral problem. This result throws light on to the algorithmic issues raised by Professor Stoyan and Professor Ziezold.

Theorem (linear–spectral). Let X_i^*: $q \times p$ denote the carrier matrix constructed from the landmark co-ordinates of the ith figure, let β_i: $p \times r$ denote the nuisance parameters, let θ: $q \times sr$ denote the mean parameters, and let ξ_i: $s \times 1$ be the covariate vector (all quantities real or complex). Then a basis for the solution of

$$\theta = \arg\left[\min_{\{\beta_i\},\,\theta} \left\{ \sum_i \| X_i^* \beta_i - \theta(\xi_i \otimes I_p) \|^2 \right\}\right] \qquad \text{such that } \theta^{\mathrm{T}}\theta = I \tag{5}$$

is the eigenvectors corresponding to the largest r eigenvalues of $\Sigma_i (H_i \otimes \xi_i \xi_i^{\mathrm{T}})$, where $H_i = X_i^* (X_i^{*\mathrm{T}} X_i^*)^{-1} X_i^{*\mathrm{T}}$ is the hat matrix.

This theorem covers the complex case (Kent), planar shapes in real co-ordinates $(q, p, r, s) = (2N, 4, 1, 1)$, affine superimposition of figures in $\mathbf{R}^K$, $(q, p, r, s) = (N, K+1, K, 1)$ and shape and affine regression $s > 1$. Constraints of the form $\mu^{\mathrm{T}}\delta$ symmetric (Section 16.1) are not enforced directly. Excluded cases are the size-and-shape problem when $K = 2$ and the size-and-shape or shape problem for $K \geqslant 3$. In the weighted least squares version, a general $q \times q$ covariance Σ is inserted in equation (5), and the H_i modified accordingly. For shapes of planar configurations, the complex case is more restrictive, as Σ is $N \times N$ complex (Kent) or $2N \times 2N$ real.

Procrustes models and identifiability

There are four steps to the Procrustes modelling approach.

(a) Define the analogous multivariate model.
(b) Include nuisance parameters for similarity transformations.
(c) Resolve questions of identifiability.
(d) Develop inference techniques in the appropriate tangent spaces.

For the one-sample model, steps (a) and (b) yield equation (5.1). The likelihood includes ridges, and for identifiability (step (c)) μ may be standardized to centroid 0, size 1, and lower triangular pattern. The pairwise model discussed by Mr Walder is more complicated. Adapting Goodall and Bose (1987), equation (11), let (X_i, Y_i), $i = 1, \ldots, L$, be the paired observations. Steps (a) and (b) give

$$\begin{aligned} X_i &= \alpha_{X_i}(\mu_i + E_{X_i})\Omega_{X_i} + \mathbf{1}_N \omega_{X_i}^{\mathrm{T}} \\ Y_i &= \alpha_{Y_i}(\mu_i + \delta + E_{Y_i})\Omega_{Y_i} + \mathbf{1}_N \omega_{Y_i}^{\mathrm{T}}. \end{aligned} \tag{6}$$

We assume that the E_{X_i} and E_{Y_i} are mutually independent. When the μ_i are random, equations (6) are comparable with Walder's set-up. For fixed δ the shape $[\mu_i + \delta]$ varies according to the choice of μ_i along the fibre of $[\mu_i]$. Conditions for identifiability are that each μ_i, and also δ, is in the Procrustes tangent space of $\bar{\mu}$, where $[\bar{\mu}]$ is the mean shape of the $[\mu_i]$ (and $\bar{\mu}$ may be standardized). This *defines* a common alignment, between the X_i and Y_i, and the rotation R_ψ in Walder's equation is unnecessary. (For small overall shape differences R_ψ is unnecessary in Walder's set-up also.)

Estimation is straightforward: superimpose the X_i by generalized Procrustes analysis, then superimpose the Y_i on the X_i' in L separate ordinary Procrustes analysis steps. To test $\delta = 0$ we compare $\hat{\delta} = \Sigma(Y_i' - X_i')/L$ with the $\{Y_i' - X_i'\}$ by a rank deficient Hotelling's T^2-statistic. This test is valid provided that the Procrustes tangent spaces $\{T_{X_i'}\}$ almost coincide.

Dr Lele points to some difficulties in estimation with many nuisance parameters. In the one-sample problem, each residual R_i is in $T_{\hat{\mu}}$, and the estimated covariance $\hat{\Sigma}$ represents pure variation in shape, between fibres (Sections 5 and 10). For example, the spectrum of the unrestricted covariance includes $K(K+1)/2$ zeros, corresponding to linear translation and rotation constraints, and, for small shape differences, one small eigenvalue corresponding to the non-linear scale constraint (equations (3.1)

and (3.2)). Thus for iteratively reweighted least squares estimation the covariance must be restricted, e.g. to $\Sigma_N = I$, as in Section 14.4. The estimate $\hat{\Sigma}_N$ (cf. equation (10.3)) includes one linear (translation) and $K(K-1)/2+1$ non-linear constraints. In this case the figures may be centred once and for all, or the empirical covariance may be shrunk towards I_N or an alternative (below) that is more realistic.

Models for variation in shape

The discussions of Professor Mardia, Professor Kent and Dr Dryden reinforce a conclusion from Section 14 that in estimating mean shape the choice of covariance does not much matter. The metric *is* important in comparing distances between shapes, e.g. in inference (Section 9). Many of the discussants point out that intrapopulation variation in shape is poorly modelled by an isotropic covariance; correlations between landmarks tend to decrease with distance. It is easy for Procrustes estimation and inference to accommodate $\Sigma_N \propto D^{-1}$. Professor Mardia's proposes a circular bivariate AR(p) error structure. The pattern of bones in Fig. 1 is more complicated than a circle, motivating me to attempt covariance selection (Dempster, 1972) to model landmark variation; the true covariance reflects various localized biological processes.

The Gaussian model is motivated by measurement error; when variation in shape *per se* predominates a probability model directly in shape space may be more attractive. Dr Dryden compares the marginal shape density with the profile likelihoods in the Procrustes approach. Despite the advantages of a marginal approach in removing the effect of nuisance parameters (Kalbfleisch and Sprott, 1970; Cox and Reid, 1987), and despite Dr Small's interesting proposal to use saddlepoint techniques to relieve some of the computational burden (and so make more general densities tractable), the simplicity of the profile likelihood (Dryden's equation (1)) suggests this to be the 'canonical' shape density, analogous to the Fisher density on the sphere. This is also an argument in favour of Procrustes methods! When $K \geqslant 3$, shape space is no longer homogeneous, and the canonical density is not a function of α alone (Goodall and Mardia, 1990a).

I am grateful to Professor Mardia for querying some aspects of my comparison of the full and marginal likelihood approaches. Development of the Bayesian approach implied there is outside the scope of this rejoinder, and the paper has been edited in proof. (I have also corrected the reference to W. S. Kendall.)

The various shape distributions are equivalent to first order (Professor Kent), but the arithmetic mean of Bookstein's variables differs from the other estimates (Professor Stoyan, Professor Ziezold and Dr Dryden). Each estimate is approximately an arithmetic mean in a tangent space, which for Bookstein's variables is fixed, and for the other estimates is tangent approximately at the mean shape. For planar triangles, Bookstein's shape space is tangent (via the stereographic projection) at the shape of equispaced collinear landmarks, and that arithmetic mean is biased in general. This limits the claim of consistency in Section 9; I am grateful to Dr Lele (personal communication) for questioning my original observation that the superimposition estimators are generally consistent (notwithstanding Professor Kent's result).

Analysis of shape differences

A theme of the discussion is that shape differences can be approached in many different ways, via

(a) informative superimpositions,
(b) transformations,
(c) the geometry of shape space and
(d) distances.

Informative superimpositions. For Dr Lele, citing Goodall and Bose (1987), the non-uniqueness of superimposition is a drawback. Any single superimposition of X and Y contains all the information about the shape difference between $[X]$ and $[Y]$, if we are sufficiently perceptive (which we are not). For the hydrocephaly data there are three representations of the difference in mean shape shown in Fig. 4 (but just one *inference*), but these are too few, and not too many, as additional superimpositions may highlight other aspects of the shape difference. The difference $Y - X'$ *must* be referenced to the corresponding X' or Y (Section 3). For linear figures, vectors X and Y, superimposition is equivalent to fitting a straight line, with $X' = \hat{Y}$. We often consider alternative fits (ordinary least squares (OLS), resistant line, . . .). The residual vector is a useful descriptor of the non-linearity, or shape difference, between X and Y, but only when it is compared with X or X'.

Transformations. The use of polynomial-based parametric transformations to describe a shape difference, due to Sneath (1967) and discussed in Section 15, is gracefully extended through Professor Bookstein's

spline-based decomposition, into principal warps (Bookstein, 1989) with the linear deformation as lowest order term. Some properties of the bending energy metric parallel those of $\Sigma \propto D^{-1}$. Shape differences in figures comprising landmarks points are usefully described as deformations of the space in which they are embedded. The deformation in the spaces between the landmarks may be interpreted physically (e.g. as a best guess at the biological growth there), but that is strictly voluntary.

In this paper equivalence classes of figures are defined by similarity transformations or rigid body motions. Equivalence classes modulo affine transformations are noted in Section 15 (Goodall and Mardia (1990a) derive the marginal density), and in the discussions of Dr Hastie and Dr Kishon, and Dr Kendall and Dr Le. Professor Mardia's example from on-line gesture recognition indicates the need for a still larger group and suggests connections to the field of pattern recognition. One further choice is projective transformations, which (in joint research) I am using to register X-ray images in a designed experiment that, incidentally, has many of the features sought by Dr Lange.

Geometry of shape space. In Professor Kendall's spherical blackboard (Kendall, 1984), and now his present discussion, each shape is shown as a point in a low dimensional view of shape space, with icons of shapes included to orient the viewer. The symmetries of shape spaces are used to focus our attention on features that are most relevant to the hypotheses. I am impressed: this is 'envisioning information' in a particularly refined sense.

The geometrical assumptions underlying the regression and growth curves techniques of Goodall (1990) and Goodall *et al.* (1990) are criticized by Professor Bookstein. He objects that the Procrustes metric is inappropriate, and biological organisms do not develop along geodesics in shape space. There is little reason to suppose that the organism grows along any other abstract trajectory, as, particularly in early development, different regions grow most rapidly at different times. Two alternatives are either to use the bending energy metric (as Bookstein proposes) or to endow shape spaces with the weighted Procrustes metric induced by general Σ. When $\Sigma = \Sigma_N \otimes I_K$ a quick solution is to premultiply each figure by Q ($Q^TQ = \Sigma_N^{-1}$, Section 7). If $\Sigma_N \propto D^{-1}$ (say), which depends on X, then Q may be a compromise 'median' choice.

Some digitizations of the rat skull include 20 landmarks, others 17, and others the neural or the facial skull alone. This realization led me to investigate the trajectories of subsets of landmarks when all N landmarks follow a geodesic. The regression models are closed under subsetting when the condition $\mu^T\delta$ or $\mu^T\mu^*$ symmetric is dropped (Section 16), and thus the shapes of triangles follow small circles. I look forward to further reports from Dr Kendall and Dr Le.

To demonstrate growth curves regression, Goodall *et al.* (1990) chose a triangle with an exceptionally large shape difference, about 16° of arc in Σ_2^3. This example shows that geometry does make a difference, seen in the different shapes traversed by the OLS great circle in shape space and the OLS straight line in the plane with Bookstein co-ordinates.

Distances. The limit of subset analysis is to consider interpoint distances. Euclidean distance matrices, multidimensional scaling and Procrustes methods appear together frequently in the research literature (e.g. in my basic references Sibson (1978, 1979)!). I am very interested to see the more specific focus of Professor Gower and Dr Critchley on shape theory and wonder what will be made of the further decomposition of the already minimal Procrustes sum of squares in the alternative geometries of E and D, or of the notions of size-shape and angular shape. Recent practical application in morphometrics is provided in the papers cited by Dr Lele. Set against the invariance of each distance is

(a) the redundancy of $\binom{N}{2}$ distances, instead of m shape variables or NK co-ordinates, and
(b) the loss of direct information about the orientations and positions of the edges, which discourages description and inferences related to transformations.

The solution is to take *more* distances, namely the space of size variables comprising distances between pairs of all possible derived landmarks (Bookstein, 1986). Then (for example) there is a derived edge parallel to the direction of maximal extension (see Fig. 6).

The use of interpoint distances by Dr Bacon-Shone in gait analysis is quite interesting. I wonder how that special geometry might be described. The compositional data approach, applied to morphometrics by Campbell and Mosimann (1987) using the Dirichlet distribution, requires a linearization of spatial data. This limitation is overcome when the underlying geometry is changed from real to complex projective space.

It is no surprise that there are so many alternatives to analyse shape differences, with some better suited to certain applications than others. Thus Dr Critchley's question (a) has more than one answer.

I have decomposed the test statistic using polynomial transformations (see also Bookstein and Sampson (1990)) and now wish to see its decomposition in distance geometry!

Closing remarks

At present the analysis of outline data is less developed than landmark methods. Professor Kent, and Dr Hastie and Dr Kishon offer some suggestions. The problem of registering polar wander paths posed by Professor Lewis is a transitional example, as the mean path may be piecewise linear. His Fig. 13 appears to be a very good superimposition already. One suggestion is to combine a combinatorial approach and spherical regression (Chang (1986), or this paper), to minimize, over $SO(3)$, either

(a) the sum of perpendicular distances between each set of points and the other piecewise linear path or
(b) the length of the piecewise linear path through the pooled data that is faithful to the partial ordering.

The (N, K, L) notation in this paper is taken from Sibson (1978, 1979) and Langron and Collins (1985), and is also used by Goodall and Mardia (1990a, b) and Carne (1990). Professor Kendall and others use the triple (k, m, n). Bookstein (1991) and Dryden and Mardia (1989a, b) use still other notation. The inconsistency is unfortunate, but I do like a shape data set to comprise n observations of a $(k \times m$?) matrix of variables.

Echoing Professor Mardia, I look forward to seeing more applications to three dimensions. Both geometry and statistics are important parts of the statistical analysis of shape. I agree with Dr Lange that more sophisticated statistical methods are worthwhile, but provided that the geometry is not oversimplified in implementing them. Professor Stoyan and Professor Ziezold mention using Procrustes techniques in cluster analysis. The graphical data analysis strategy of Weihs and Schmidli (1990) shows that Procrustes analysis, including inference, can have very general application.

REFERENCES IN THE DISCUSSION

Aitchison, J. (1986) *The Statistical Analysis of Compositional Data*. London: Chapman and Hall.

Ambartzumian, R. V. (1990) *Factorization Calculus and Geometric Probability*. Cambridge: Cambridge University Press.

Bookstein, F. L. (1986) Size and shape spaces for landmark data in two dimensions (with discussion) *Statist. Sci.*, **1**, 181–242.

——— (1989) Comment on A survey of the statistical theory of shape (by D. G. Kendall). *Statist. Sci.*, **4**, 99–105.

——— (1991) *Morphometric Tools for Landmark Data: Geometry and Biology*. Cambridge: Cambridge University Press. To be published.

Bookstein, F. L. and Sampson, P. D. (1990) Statistical models for geometric components of shape change. *Communs Statist. Theory Meth.*, **19**, 1939–1972.

Broadbent, S. (1980) Simulating the ley hunter (with discussion). *J. R. Statist. Soc.* A, **143**, 109–140.

Campbell, G. and Mosimann, J. E. (1987) Multivariate methods for proportional shape. *Proc. Statist. Graphics Sect. Am. Statist. Ass.*, 10–17.

Carne, T. K. (1990) The geometry of shape spaces. *Proc. Lond. Math. Soc.*, **61**, 407–432.

Carroll, J. D. and Chang, J. J. (1970) Analysis of individual differences in multidimensional scaling via an n-way generalization of "Eckart–Young" decomposition. *Psychometrika*, **35**, 283–319.

Chang, T. (1986) Spherical regression. *Ann. Statist.*, **14**, 907–924.

Chow, Y., Grenander, U. and Keenan, D. M. (1988) HANDS, a pattern theoretic study of biological shapes. *Technical Report*. Division of Applied Mathematics, Brown University, Providence.

Cox, D. R. and Reid, N. (1987) Parameter orthogonality and approximate conditional inference (with discussion). *J. R. Statist. Soc.* B, **49**, 1–39.

Critchley, F. (1980) Optimal norm characterisations of multidimensional scaling methods and some related data analysis problems. In *Data Analysis and Informatics* (eds E. Diday *et al.*), pp. 209–229. Amsterdam: North-Holland.

——— (1986) Dimensionality theorems in multidimensional scaling and hierarchical cluster analysis. In *Data Analysis and Informatics, IV* (eds E. Diday *et al.*), pp. 45–70. Amsterdam: North-Holland.

——— (1988) On certain linear mappings between inner-product and squared-distance matrices. *Lin. Alg. Applicns*, **105**, 91–107.

——— (1991) On certain linear transformations of the cone of symmetric nonnegative definite matrices. *Warwick Statistics Research Report*. University of Warwick, Coventry.

Dempster, A. P. (1972) Covariance selection. *Biometrics*, **28**, 157–175.

Dryden, I. L. and Mardia, K. V. (1991) General shape distributions in the plane. *Adv. Appl. Probab.*, to be published.

Embleton, B. J. J., Fisher, N. I. and Schmidt, P. W. (1983) Analytic comparison of apparent polar wander paths. *Earth Planet. Sci. Lett.*, **64**, 276–282.

Goodall, C. R. (1990) Regression models for shape and form. To be published.

Goodall, C. R. and Bose, A. (1987) Procrustes techniques for the analysis of shape and shape change. In *Computer Science and Statistics* (ed. R. M. Heiberger), pp. 86–92. Alexandria: American Statistical Association.
Goodall, C. R., Lange, N. and Moss, M. L. (1990) Growth-curve models for repeated triangular shapes. To be published.
Goodall, C. R. and Mardia, K. V. (1990a) Multivariate aspects of shape theory. Submitted to *Ann. Statist.*
——— (1990b) A geometrical derivation of the shape density. *Adv. Appl. Probab.*, to be published.
Gough, J. and Mardia, K. V. (1990) Shape modelling using deformable polygons. *Technical Report*. Department of Statistics, University of Leeds.
Gower, J. C. (1970) Statistical methods of comparing different multivariate analyses of the same data. In *Mathematics in the Archaeological and Historical Sciences*. Edinburgh: Edinburgh University Press.
——— (1975) Generalized Procrustes analysis. *Psychometrika*, **40**, 33–50.
——— (1985) Properties of Euclidean and non-Euclidean distance matrices. *Lin. Alg. Applicns*, **67**, 81–97.
Kalbfleisch, J. D. and Sprott, D. A. (1970) Application of likelihood methods to problems involving large numbers of parameters (with discussion). *J. R. Statist. Soc.* B, **32**, 175–208.
Kendall, D. G. (1984) Shape-manifolds, procrustean metrics and complex projective spaces. *Bull. Lond. Math. Soc.*, **16**, 81–121.
Kendall, D. G. and Kendall, W. S. (1980) Alignments in two-dimensional random sets of points. *Adv. Appl. Probab.*, **12**, 380–424.
Kendall, W. S. (1990a) The diffusion of Euclidean shape. In *Disorder in Physical Systems* (eds D. Welsh and G. Grimmett), pp. 203–217. Oxford: Oxford University Press.
——— (1990b) Probability, convexity, and harmonic maps with small image: I, Uniqueness and fine existence. *Proc. Lond. Math. Soc.*, **61**, 371–406.
Lange, N., Carlin, B. P. and Gelfand, A. E. (1990) Hierarchical Bayes models for longitudinal CD4$^+$ counts. *Technical Report*. Division of Biology and Medicine, Brown University, Providence.
Langron, S. P. and Collins, A. J. (1985) Perturbation theory for generalized Procrustes analysis. *J. R. Statist. Soc.* B, **47**, 277–284.
Lele, S. R. (1991) Some comments on coordinate free and scale invariant methods in morphometrics. *Am. J. Phys. Anthrop.*, to be published.
Lele, S. R. and Richtsmeier, J. T. (1991a) A coordinate free approach for comparing biological shapes: landmark data. *Am. J. Phys. Anthrop.*, to be published.
——— (1991b) On comparing biological shapes: detection of influential landmarks. *Am. J. Phys. Anthrop.*, to be published.
Mardia, K. V. (1984) Spatial discrimination and classification maps. *Communs Statist. Theory Meth.*, **13**, 2181–2197.
——— (1989) Shape analysis of triangles through directional techniques. *J. R. Statist. Soc.* B, **51**, 449–458.
Mardia, K. V. and Dryden, I. L. (1989a) The statistical analysis of shape data. *Biometrika*, **76**, 271–282.
——— (1989b) Shape distributions for landmark data. *Adv. Appl. Probab.*, **21**, 742–755.
Mardia, K. V., Sheehy, N., Ghali, N. and Hainsworth, T. (1990) *Techniques for On-line Character Recognition*. New York: Institute of Electronic and Electrical Engineers. To be published.
Neyman, J. and Scott, E. (1948) Consistent estimates based on partially consistent observations. *Econometrics*, **16**, 1–32.
Rice, J. and Silverman, B. (1991) Estimating the mean and covariance structure nonparametrically when the data are curves. *J. R. Statist. Soc.* B, **53**, 233–243.
Richtsmeier, J. T. and Lele, S. R. (1990) Analysis of craniofacial growth in Crouzon syndrome using landmark data. *J. Cranfac. Genet. Devlpmnt. Biol.*, **10**, 39–62.
Rohlf, F. J. and Slice, D. (1990) Methods for comparison of sets of landmarks. *Syst. Zool.*, **39**, 40–59.
Sibson, R. (1978) Studies in the robustness of multidimensional scaling: Procrustes statistics. *J. R. Statist. Soc.* B, **40**, 234–238.
——— (1979) Studies in the robustness of multidimensional scaling: perturbational analysis of classical scaling. *J. R. Statist. Soc.* B, **41**, 217–229.
Sneath, P. H. A. (1967) Trend-surface analysis of transformation grids. *J. Zool.*, **151**, 65–122.
Stoyan, D. (1990) Estimation of distances and variances in Bookstein's landmark model. *Biometr. J.*, **32**, 843–849.
Stoyan, D. and Stoyan, H. (1990) A further application of D. G. Kendall's procrustes analysis. *Biometr. J.*, **32**, 293–301.
Weihs, C. and Schmidli, H. (1990) OMEGA (online multivariate exploratory graphical analysis): routine searching for structure (with discussion). *Statist. Sci.*, **5**, 175–226.
Ziezold, H. (1977) On expected figures and a strong law of large numbers for random elements in quasi-metric spaces. In *Trans. 7th Prague Conf. Inference Theory, Statistics of Decision Functions, Random Processes, Prague, 1974*, vol. A, pp. 591–602. Dordrecht: Reidel.
——— (1989) On expected figures in the plane. *Math. Res. Ser.*, **51**, 105–110.